

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**  
**МІСЬКОГО ГОСПОДАРСТВА імені О. М. БЕКЕТОВА**

**О. О. ВОРОНКОВ**

КОНСПЕКТ ЛЕКЦІЙ  
з курсу

**ЕКОНОМІКО-МАТЕМАТИЧНЕ**  
**МОДЕЛЮВАННЯ**

*(для студентів заочної форми навчання освітньо-кваліфікаційного рівня  
«бакалавр» напрямку підготовки 6.030601 - Менеджмент)*

**Харків**  
**ХНУМГ ім. О. М. Бекетова**  
**2015**

**Воронков О. О.** Конспект лекцій з курсу «Економіко-математичне моделювання» (для студентів заочної форми навчання освітньо-кваліфікаційного рівня «бакалавр» напряму підготовки 6.030601 - Менеджмент) / О. О. Воронков; Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. – Харків: ХНУМГ ім. О. М. Бекетова, 2015. – 148 с.

Автор: канд. екон. наук О. О. Воронков

Рецензент: канд. екон. наук, доцент кафедри ЕПМГ Г. І. Базецька

*Рекомендовано кафедрою економіки підприємств міського господарства,  
протокол № 1 від 27.08.2014 р.*

## ВСТУП

Курс «Економіко-математичне моделювання» є нормативною дисципліною циклу природничо-наукової і загальноекономічної підготовки в навчальному плані за напрямом «Менеджмент» для кваліфікаційного рівня «Бакалавр». Обсяг курсу становить 144 академічних години або 4 кредити. При вивченні за заочною формою обсяг аудиторних занять становить 16 годин (10 годин лекцій і 6 годин практичних занять), на самостійну роботу студента припадає 128 годин. Програма курсу містить 2 змістових модулів: «Оптимізаційні економіко-математичні моделі» і «Економетричні моделі», відповідно до яких виконується проміжний контроль знань. Підсумковий контроль знань (іспит) проводиться в усній формі. У процесі вивчення курсу студенти повинні виконати розрахунково-графічну роботу.

Метою вивчення дисципліни «Економіко-математичне моделювання» є формування системи базових знань в області методології постановки задач, побудови економіко-математичних моделей та методів їхнього розв'язання і аналізу.

В результаті вивчення курсу студенти повинні оволодіти прийомами побудови економіко-математичних моделей, основними математичними поняттями і методами розв'язання оптимізаційних задач різної складності, а також користатися методами економетричного моделювання.

Сучасна економічна наука характеризується широким використанням математики. Математичні методи є складовою частиною методів будь-якої економічної науки, включаючи економічну теорію. Використання математичних методів відчиняє нові можливості, і фахівцеві необхідно вміти формулювати та вирішувати задачі оптимізації виробництва, моделювати економічну динаміку і ризикові ситуації, статистично оцінювати економічні залежності, а також користуватися ігровими методами.

Математичні моделі використовувалися з ілюстративними та дослідницькими цілями ще Ф.Кене (1758 р., «Економічна таблиця»), А.Смітом (класична макроекономічна модель), Д.Рикардо (модель міжнародної торгівлі). В XIX столітті великий внесок у моделювання ринкової економіки внесла математична школа (Л.Вальрас, О.Курно, В.Парето, Ф.Еджворт та ін.). У XX столітті математичні методи моделювання застосовувалися дуже широко, з їхнім використанням пов'язані практично всі роботи, що відзначені Нобелівською премією з економіки (Д.Хікс, Р.Солоу, В.Леонт'єв, П.Самуельсон та ін.). Розвиток мікроекономіки, макроекономіки, прикладних дисциплін пов'язаний з усе більш високим рівнем їхньої формалізації. Основу для цього заклав прогрес в галузі прикладної математики - теорії ігор, математичного програмування, математичної статистики.

Перші роботи з економетрики з'явилися наприкінці XIX – початку XX століття. У 1897 р. з'явилася робота одного із засновників математичної школи в економічній теорії В. Парето, що була присвячена статистичному вивченню доходів населення в різних країнах. Запропонована крива Парето  $y = A(x-a)^{-\alpha}$ , де  $x$  - величина доходу;  $y$  - чисельність осіб, що мають дохід, що перевищує  $x$ ;

$a$  - мінімальний дохід;  $A$  і  $\alpha$  - параметри залежності, одержувані за статистичними методами.

На самому початку XX століття вийшло кілька робіт англійського статистика Гукера, у яких він застосував кореляційно-регресійні методи, розроблені Пірсоном та його школою, для вивчення взаємозв'язків економічних показників, зокрема - впливу числа банкрутств на товарній біржі на ціну зерна. У роботах Гукера містилася ідея часового лага між економічними змінними, а також ідея кореляційного аналізу не самих величин, а їхніх приростів. Надалі з'явилося величезна кількість робіт як з розвитку теорії математичної статистики та її прикладних елементів, так і з практичного застосування цих методів в економічному аналізі. До першої групи можуть бути, наприклад, віднесені роботи Р.Фішера з дисперсійного аналізу, до другого - роботи з оцінки й дослідження виробничих функцій, зокрема класична робота Кобба й Дугласа 1928 р. Економетричні моделі й методи зараз - це не тільки могутній інструментарій для одержання нових знань в економіці, але й широко застосовуваний апарат для прийняття практичних рішень у прогнозуванні, банківській справі, бізнесі.

### **ЗМ 1 Оптимізаційні економіко-математичні моделі**

#### **ТЕМА 1**

### **КОНЦЕПТУАЛЬНІ АСПЕКТИ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ ЕКОНОМІКИ**

#### **1.1 Економічні моделі. Поняття економічної моделі**

Сучасна економічна теорія, як на мікро-, так і на макрорівні, включає як необхідний елемент математичні моделі і методи. Використання математики в економіці дозволяє, по-перше, виділити і формально описати найбільш важливі, істотні зв'язки економічних змінних і об'єктів. По-друге, із чітко сформульованих вихідних даних і співвідношень методами дедукції можна одержувати висновки, адекватні досліджуваному об'єкту. По-третє, методи математики і статистики дозволяють індуктивним шляхом одержувати нові знання про об'єкт: оцінювати форму і параметри залежностей його змінних, що найбільшою мірою відповідають наявним спостереженням. І, нарешті, по-четверте, використання мови математики дозволяє точно і компактно викладати положення економічної теорії, формулювати її поняття і висновки.

Для вивчення різних економічних явищ економісти використовують їх спрощені формальні описи, називані **економічними моделями**. Прикладами економічних моделей є моделі споживчого вибору, моделі фірми, моделі економічного зростання, моделі рівноваги на товарних, факторних і фінансових ринках та багато інших. При побудові моделі виявляють істотні фактори, що зумовлюють досліджуване явище і відкидають деталі, неістотні для вирішення поставленої проблеми. Формалізація основних особливостей функціонування

економічних об'єктів дозволяє оцінити можливі наслідки впливу на них і використовувати такі оцінки в управлінні.

Модель (лат. - міра, зразок) - це певний об'єкт, який у певних умовах заміняє об'єкт-оригінал. Модель відтворює властивості оригіналу, що підлягають дослідженню, та його характеристики. Модель має при цьому істотні переваги під час використання, зокрема, наочність, доступність випробувань та ін.

Моделювання - це дослідження будь-яких процесів, явищ або систем (об'єктів) шляхом побудови й вивчення їхніх моделей; використання моделей для визначення або уточнення характеристик і раціоналізації способів побудови новостворюваних об'єктів.

Моделювання - одна з основних категорій теорії пізнання. На ідеї моделювання базується будь-який метод наукового дослідження, як теоретичний, при якому використовуються абстрактні моделі, так і експериментальний, що використовує предметні моделі.

## 1.2 Принципи моделювання

Побудова і використання моделей базується на наступних основних принципах.

1. Принцип інформаційної достатності. При повній відсутності інформації про систему побудова її моделі є неможливою. За наявності повної інформації про систему відсутня необхідність її моделювання. У всіх проміжних ситуаціях дослідник змушений оперувати з моделлю системи, рівень адекватності якої зумовлюється деяким критичним рівнем апріорної інформації про систему.

2. Принцип здійсненості. Модель, розроблювана дослідником, має забезпечити досягнення поставленої мети із практичною вірогідністю і за кінцевий час.

3. Принцип множинності моделей. Складність досліджуваних систем і велика розмаїтість їхніх властивостей не дозволяють побудувати одну досить адекватну модель. Тому виникає необхідність у побудові кількох моделей, які в сукупності дають досить повне уявлення про систему і всі процеси, що відбуваються у ній.

Будь-яка діяльність людини носить **цільовий характер**. Моделювання є неминучою процедурою у всякій доцільній діяльності. Треба особливо підкреслити, що модель є не взагалі якимсь відображенням оригіналу, а цільовим відображенням. Ціль моделювання зумовлює, які властивості оригіналу і якою мірою (з якою точністю) повинна відображати модель.

Як зазвичай будують економічну модель?

1. Формулюють предмет і цілі дослідження.

2. У розглянутій економічній системі виділяють структурні або функціональні елементи, що відповідають даній цілі, виявляють найбільш важливі якісні характеристики цих елементів.

3. Словесно, якісно описують взаємозв'язки між елементами моделі.

4. Вводять символічні позначення для характеристик економічного об'єкта, що враховуються, і формалізують, наскільки можливо, взаємозв'язки між ними. Тим самим, формулюється математична модель.

5. Проводять розрахунки за математичною моделлю і аналіз отриманого розв'язку.

Розглянемо математичну структуру моделі та її змістовну інтерпретацію.

**Приклад 1.** Нехай потрібно визначити, яку суму треба покласти в банк при заданій ставці відсотку (20% річних), щоб через рік одержати \$12000. Введемо формальні позначення для величин: початкова сума грошей -  $M_0$ ; кінцева сума грошей -  $M_1$ ; ставка відсотка -  $R$ . Запишемо співвідношення між ними

$$M_1 = M_0 \left( 1 + \frac{R}{100} \right)$$

та знайдемо необхідну величину з розв'язання основного рівняння моделі

$$M_0 = \frac{M_1}{1 + \frac{R}{100}} = \frac{\$12000}{1,2} = \$10000.$$

**Приклад 2.** Нехай потрібно визначити, який був обсяг випуску продукції заводу, якщо в результаті технічного переозброєння середня продуктивність праці збільшилася на 20%, і завод став випускати 12000 одиниць продукції. Введемо формальні позначення для величин: початковий випуск -  $Q_0$ ; кінцевий випуск -  $Q_1$ ; відсоток приросту продуктивності -  $R$  і запишемо співвідношення між ними (що випливає з визначення середньої продуктивності праці  $\frac{Q}{L}$ )

$$Q_1 = Q_0 \frac{L_1}{L_0} = Q_0 \left( 1 + \frac{L_1 - L_0}{L_0} \right) = Q_0 \left( 1 + \frac{R}{100} \right).$$

Знайдемо шукану величину з розв'язання основного рівняння моделі:

$$Q_0 = \frac{Q_1}{1 + \frac{R}{100}} = \frac{\$12000}{1,2} = \$10000.$$

Порівнюючи отримані моделі і результати, можемо помітити, що математична форма моделі

$$X_1 = X_0 \left( 1 + \frac{R}{100} \right)$$

і навіть числові значення величин, що входять до неї, в обох випадках однакові, але економічна ситуація, описувана моделлю, економічна інтерпретація моделі і результатів розрахунку зовсім різні. Таким чином, ті самі математичні моделі і методи можуть бути використані для вирішення зовсім різних економічних завдань.

### 1.3 Класифікація моделей

Підставою для класифікації моделей є ціль моделювання. Цільова орієнтація моделей дозволяє класифікувати їх за типами цілей, за способами відтворення (реалізації), за зміною на різних етапах життєвого циклу.

За типами цілей розрізняють моделі пізнавальні і прагматичні. За способами відтворення розрізняють моделі ідеальні (абстрактні) і матеріальні (реальні, фізичні). За зміною в часі розрізняють моделі статичні і динамічні.

Розходження пізнавальних і прагматичних моделей проявляється в їхньому ставленні до оригіналу в процесі діяльності. **Пізнавальні моделі** є формою організації і подання знань, засобами з'єднання нових знань із наявними. Якщо в процесі створення пізнавальної моделі деякого реального об'єкта спостерігаються розбіжності, то здійснюється корекція моделі з метою наблизити її до реальності.

**Прагматичні моделі** є засобами управління практичними діями, способом уявлення необхідних дій або їхнього результату.

Прагматичні і пізнавальні моделі можуть створюватися як для деякого фіксованого моменту часу, тобто бути статичними, так і для деякого інтервалу життєвого циклу об'єкта в умовах середовища, що змінюється. У цьому випадку необхідні динамічна модель, параметри якої залежать від часу. Динамічні моделі зазвичай значно складніше за статичні, вимагають застосування складного математичного апарата і великого обсягу машинної пам'яті. Для спрощення динамічні моделі часто розглядають у вигляді набору послідовностей статичних моделей, що відповідають певним моментам часу на траєкторії життєвого циклу об'єкта.

**Ідеальними** називають моделі, що побудовані засобами мислення, свідомості. До цих моделей належать всі мовні конструкції, що сприяють встановленню відносин між людьми.

Природна мова є універсальним засобом побудови абстрактних конструкцій (моделей). Ця універсальність забезпечується можливістю введення в мову нових слів і можливістю ієрархічної побудови усе більш розвинених мовних моделей. Крім цього, універсальність мови досягається тим, що мовні конструкції мають неоднозначність, розпливчастість, розмитість. Людина переборює цю неоднозначність за допомогою «розуміння» і «інтерпретації», що залежить від культурного середовища, у якому вона живе.

Якщо на побутовому рівні приблизність мови сприяє поліпшенню відносин серед людей, то при вирішенні ряду практичних і наукових завдань це стає перешкодою, наприклад, у діалозі з ЕОМ. Тому з'являється «професійна мова», реалізована в обмежених професійних колективах, мова математиків, фізиків, медиків, льотчиків, фахівців з інформатики та ін.

Найбільш високоспеціалізованою є мова математики, що має максимально досягну на сьогодні визначеність і точність.

Еммануїл Кант відзначав, що «у кожному пізнанні стільки науки, скільки є в ньому математики». Низький рівень математизації будь-якої науки не означає її «ненауковість», а просто є наслідком її складності, недостатньої

глибини пізнання предмета науки й, отже, явищем тимчасовим. Наочним прикладом цьому є зміна поглядів людей на проблему створення «мислячої машини». Якщо в 50-х роках нашого століття розмови про це викликали тільки посмішки скептиків і гнів офіційної науки, то зараз з'явився цілий напрямок у математиці - теорія штучного інтелекту, на підставі якої базуються високоефективні експертні системи.

За характером обчислення різних показників математичні моделі діляться на аналітичні, алгоритмічні та імітаційні.

**Аналітичні моделі** передбачають реалізацію моделі у вигляді алгебраїчних, диференціальних, інтегральних та інших рівнянь, що зв'язують вихідні змінні із вхідними доповненою системою обмежень. При цьому передбачається наявність однозначної обчислювальної процедури одержання точного розв'язку рівнянь.

При алгоритмічному підході використовується математична модель не допускає точного подання у вигляді системи рівнянь, а залежність між вхідними (незалежними) і вихідними (залежними) змінними задається певною послідовністю дій. Такий підхід при створенні моделей складних систем є типовим.

Основним видом математичних моделей складних систем є імітаційні моделі.

**Імітаційна модель** являє собою певну обчислювальну процедуру, що описує об'єкт аналізу. Вона дозволяє з будь-якою заданою точністю відтворити параметричну систему будь-якої складності. Основними обмеженнями при створенні цих моделей є ресурси пам'яті й часу. Головними засобами реалізації імітаційних моделей є ЕОМ. Імітаційне моделювання, являючись своєрідним машинним експериментуванням з моделлю реальної системи, відкриває широкі можливості адекватного відображення процесу функціонування складних систем.

Математичні моделі, що використовуються в економіці, можна підрозділити на класи за рядом ознак, що належать до особливостей моделюваного об'єкта, цілі моделювання і використовуваного інструментарію: моделі макро- і мікроекономічні, теоретичні і прикладні, оптимізаційні і рівноважні, статичні і динамічні.

**Теоретичні моделі** дозволяють вивчати загальні властивості економіки і її характерних елементів дедукцією висновків з формальних передумов. **Прикладні моделі** дають можливість оцінити параметри функціонування конкретного економічного об'єкта і сформулювати рекомендації для прийняття практичних рішень. До прикладних належать насамперед економетричні моделі, що оперують числовими значеннями економічних змінних і дозволяють статистично значущо оцінювати їх на основі наявних спостережень.

У моделях **статичних** описується стан економічного об'єкта в конкретний момент або період часу; **динамічні моделі** включають взаємозв'язки змінних у часі. У статичних моделях звичайно зафіксовані значення ряду величин, що є змінними в динаміці, - наприклад, капітальних ресурсів, цін та ін. Динамічна модель не зводиться до простої суми ряду статичних, а описує сили і взаємодії в економіці, що обумовлюють хід процесів у ній. Динамічні моделі звичайно



використовують апарат диференціальних і різницевих рівнянь, а також варіаційного обчислення.

**Детерміновані моделі** припускають жорсткі функціональні зв'язки між змінними моделі. **Стохастичні моделі** припускають наявність випадкових впливів на досліджувані показники і використовують інструментарій теорії імовірностей і математичної статистики для їхнього опису.

### 1.4 Якість моделі

Якість моделі, як і будь-якого іншого об'єкта, являє собою сукупність різних властивостей, що мають ієрархічну структуру. На першому рівні виділяються три групи властивостей: цільові, експлуатаційні, модифікаційні.

**Цільовими** називаються властивості моделі, що характеризують ступінь її відповідності цілям і завданням конкретного об'єкта моделювання. Цільові властивості включають наступні властивості другого рівня: адекватність, стійкість, точність, результативність.

**Експлуатаційними** називають властивості моделі системи, що визначають зручність експлуатації моделі користувачем. До них належать наступні властивості другого рівня: продуктивність, надійність, захищеність, завершеність.

**Модифікаційними** називають властивості моделі, що обумовлюють зручності внесення зміни в модель, простоту модернізації моделі при розвитку й удосконалюванні досліджуваної системи при переході на інший рівень дослідження (наприклад, при підвищенні ступеня деталізації математичного опису). До цієї групи належать наступні властивості: зрозумілість, структурованість, розширюваність, доступність.

У людській практиці спрощеність моделей є припустимою, більше того, у деяких випадках вона є необхідною. І хоча спрощеність моделей приводить до розходження між моделлю і оригіналом, міру припустимого розходження можна оцінити тільки, якщо співвіднести її з ціллю моделювання. Тому для оцінки ступеня відповідності моделі й оригіналу вводиться таке поняття як "адекватність моделі".

Адекватною називають таку модель, для якої вимоги повноти, точності й істинності моделі виконуються не взагалі повною мірою, а тільки в тій мірі, що приводить до досягнення цілі.

### 1.5 Прийняття рішень (вибір)

Вибір є дією, що дозволяє організувати цілеспрямовану діяльність людини, а рішення - результат цього вибору у певній нормативно-правовій формі (порада, рекомендація, наказ, програма та ін.).

Управляючи будь-якою діяльністю, людина постійно зіштовхується з необхідністю приймати рішення на ту або іншу дію. Таким чином, ухвалення рішення розглядається як процедура вибору альтернативи із заданої множини.

**Альтернатива** (фр.) - кожна з взаємовиключаючих двох або більше можливостей.

Для того щоб прийняти рішення, необхідно виконати наступну послідовність процедур:

- 1) вибрати цілі, заради досягнення яких здійснюється вибір;
- 2) оцінити ступінь погодженості цілей (від повної погодженості до повного протиріччя);
- 3) сформулювати множину альтернатив, з яких здійснюється вибір;
- 4) проаналізувати і оцінити наслідки реалізації кожної альтернативи;
- 5) сформулювати критерії порівняння, тобто правило, за допомогою якого визначається перевага альтернатив;
- 6) задати режим вибору: однократний або багаторазовий з навчанням;
- 7) оцінити ситуацію, у якій здійснюється вибір (визначеність або невизначеність і її вид);
- 8) визначити тип відповідальності (індивідуальна, групова).

Сучасна наука розрізняє дві частини теорії прийняття рішень.

Одна з них має нормативний характер і займається дослідженнями технологій прийняття рішень для того, щоб можна було відповісти на запитання: «Як необхідно приймати рішення?».

Інша частина досліджує питання, пов'язані з тим, як люди на практиці приймають рішення, і які при цьому вони роблять помилки.

Нормативна теорія прийняття рішень у літературі часто називається теорією прийняття оптимального рішення, а друга частина психологічною теорією рішень.

**Теорія прийняття оптимального рішення** розвивалася завдяки успіхам, досягнутим в області дослідження операцій і системотехніці. Ця теорія розробляє методологію прийняття рішень з метою організації діяльності з досягнення певної цілі, але не дає відповідь на питання про те, як ця ціль вибирається і чи відповідає вона бажаному сценарію розвитку подій.

Вона оперує з процедурами і критеріями прийняття рішень, які можна вважати оптимальними в рамках тієї моделі розглянутої ситуації, якою керувалася особа, що приймає рішення, опираючись на свій рівень знань і наявну інформацію.

Одним з найважливіших напрямків теорії прийняття оптимальних рішень є опис умов, які повинні бути виконані для того, щоб рішення було оптимальним. Ці умови формулюються у вигляді постулатів оптимальності, серед яких найбільше визнання одержали два: постулати послідовності і максимізації.

**Постулат послідовності** говорить, що для прийняття оптимального рішення треба впорядкувати сукупність альтернатив з погляду переваги особи, що приймає рішення (ОПР). Якщо розглядаються три альтернативи  $A$ ,  $B$  і  $C$ , то впорядкування цих альтернатив означає побудову шкали строгого або слабкого порядку альтернатив за ступенем убуття їхньої переваги. Наприклад, порядок альтернатив виду  $A > C > B$  означає, що при здійсненні вибору  $A$  переважніше  $C$ , а  $C$  переважніше  $B$ .

**Постулат максимізації** затверджує, що остаточною умовою оптимального рішення є використання принципу максимізації, тобто вибір такої дії, що максимізує цільову функцію.

Формалізація моделей прийняття рішень для розглянутих ситуацій здійснюється в термінах теорії імовірностей, математичної статистики, теорії масового обслуговування, теорії; ігор та ін.

Застосування формальних оптимізаційних методів підготовки і прийняття рішень можливо тільки в умовах досить повної інформації. Така ситуація виникає, як правило, на нижніх рівнях управлінської структури, де як об'єкти управління розглядаються технічні системи або технологічні процеси.

Область застосування оптимізаційних методів сильно звужується, коли доводиться приймати рішення на верхніх рівнях управління, де об'єкти, події, явища і процеси слабо структуровані.

Психологічна теорія прийняття рішень - це система мотивованих стверджень, що розкривають внутрішній зміст діяльності людини в процесі підготовки і прийняття рішень. Практика показує, що поведінка людини в сильному ступені залежить від структури розв'язуваного завдання. Отже, одним з головних напрямків розвитку цієї теорії є класифікація завдань і опис їхніх інваріантних властивостей з погляду їхнього впливу на характер прийняття рішень різними групами людей.

У психологічній теорії прийняття рішень виділяють кілька напрямків, для яких формулюються ствердження і рекомендації.

До першого напрямку належать питання, пов'язані з моделями уявлення оцінюваної ситуації різними категоріями людей. Наприклад, психологи виявили, що більшість людей схильні до спрощення і недооцінки рівня напруженості ситуації і не дуже охоче формулюють різні альтернативи.

Другий напрямок охоплює питання, пов'язані з моделюванням процесів визначення цінності того або іншого варіанта дії, з погляду його корисності для досягнення цілі.

Третій напрямок займається дослідженням суб'єктивної оцінки імовірностей появи тих або інших факторів і їхніх рівнів, які визначають наслідки прийнятих рішень. Так, наприклад, виявлено, що люди переоцінюють імовірності рідких подій і одночасно недооцінюють міру можливості появи тих подій, об'єктивна імовірність яких велика.

Четвертий напрямок вивчає стратегії вибору поведінки людини при ухваленні рішення. Розробляються моделі, що описують, як ОПР інтерпретують інформацію про корисність наслідків і імовірностях їхнього настання і які правила вибору при цьому використовуються. Учені виявили, що в простих завданнях, пов'язаних з ризиком, люди звичайно вибирають стратегії, які максимізують суб'єктивно очікувану корисність, що вони уявляють як лінійну комбінацію суб'єктивних імовірностей наслідків і їх корисностей.

П'ятий напрямок займається описом факторів, керуючих процесом прийняття рішення. Такі фактори враховують вплив навколишнього природного середовища, особові особливості ОПР, вплив соціальної групи та ін.

Основою методології досліджень у психологічній теорії прийняття рішень є лабораторний експеримент, що включає кілька етапів.

На першому етапі формулюється сукупність аксіом і стверджень, що стосуються певних об'єктів, таких, наприклад, як ризик або перевага.

На другому етапі за допомогою формальних логічних міркувань експертів виводяться нові ствердження про можливий результат експерименту.

На третьому етапі здійснюється експериментальна перевірка висловлених гіпотез.

У цей час, поряд з натурними експериментами, для вирішення перерахованих вище завдань широко використовується машинний експеримент.

Сучасні інформаційні технології, швидкодія і великий обсяг пам'яті ЕОМ дозволяють розробляти імітаційні моделі практично будь-якої складності і одержувати необхідні рекомендації.

## **1.6 Методи прийняття рішень**

У психології прийняття людиною рішення зв'язують з процесом антиципації. Антиципація - це психічний процес, що забезпечує можливість приймати рішення з певним часово-просторовим попередженням подій, з «забіганням уперед».

У практиці людської діяльності розрізняють наступні методи прийняття рішень.

**Автоматичний метод**, реалізований на рівні біологічних механізмів інстинктів і рефлексів.

**Метод проб і помилок**, коли здійснюється несвідомий (некерований) або свідомий (керований) перебір ситуацій і пошук найкращого варіанта (режим самонавчання).

**Метод звертання до авторитетів**. Цей метод має багато прикладних відтінків: від спроби ОПР ухилитися від відповідальності (посиланням на бога, вождя, начальника та ін.) До спроби обпертися на досягнення кращих розумів людства.

**Математичний метод** є найбільш формалізованим методом прийняття рішень і вимагає наявності наступних обов'язкових компонентів:

- перелік цілей;
- перелік альтернатив;
- методи для прогнозування наслідків цих альтернатив;
- будь-який метод присвоєння імовірностей цим наслідкам (якщо вони існують);
- система цінностей, схована у цілях, для визначення цінності наслідків;
- критерій рішення, включений до системи цінностей, і який вказує, як визначити найкращу альтернативу.

Така модель прийняття рішень називається **теорією статистичних рішень**.

На жаль, на практиці лише незначна частина рішень приймається за цим методом. Це пов'язано зі слабкою структурованістю цілей, низьким рівнем формалізації альтернатив і відсутністю достовірної інформації про закони розподілу імовірностей наслідків, що розглядаються.

Залежно від типу зв'язку між альтернативами і їхніми наслідками виникають деякі умови, у яких доводиться приймати рішення. Вони визначаються як умови визначеності або невизначеності.

Ухвалення рішення в умовах визначеності характеризуються наявністю повної і достовірної інформації про проблемну ситуацію, цілі, обмеження і наслідки прийнятих рішень. У таких завданнях заздалегідь відомі наслідки всіх альтернатив.

На практиці приймати рішення доводиться в умовах, коли інформація про множину факторів, що впливають на вибір стратегії поведінки, як правило, неповна. Це приводить до того, що кожної розглянутої стратегії поведінки (альтернативи) у загальному випадку відповідає декілька можливих наслідків, що істотно ускладнює вибір стратегії.

Розрізняють стохастичну і нестохастичну невизначеності.

Стохастична невизначеність - це такий стан рівня інформованості ОПР, при якому для кожної розглянутої альтернативи відомі всі можливі наслідки й імовірності їхньої реалізації, тобто закон розподілу імовірностей наслідків. Прийняття рішення в умовах стохастичної невизначеності приводить до того, що поява того або іншого обраного результату є подією випадковою. Отже, після ухвалення рішення ОПР із тією або іншою імовірністю повинна очікувати, що відбудеться не та подія, на яку він розраховував. Завдання такого типу в теорії прийняття рішень називають **«прийняттям рішень в умовах ризику»**.

Ризик визначається як незапланований (небажаний) результат розглянутого виду діяльності організації або спосіб функціонування якогось об'єкта.

Як міра ризику розглядають або імовірність небажаного результату, або математичне сподівання міри збитку, отриманого при реалізації стратегії.

Природно, що при прийнятті рішень у таких умовах ОПР повинна прагнути мінімізувати ризик або, принаймні, контролювати його рівень. При цьому виникають різні можливості організації поведінки ОПР, тобто прийняття рішень за різними критеріями.

У тому випадку, коли інформація про закон розподілу імовірностей наслідків відсутня, ОПР змушена використовувати методи прийняття рішень, не зв'язані із стохастичною природою розглянутих процесів, наприклад, методи теорії ігор. Такий вид невизначеностей називають нестохастичною невизначеністю.

Нестохастична невизначеність підрозділяється на природну і поведінкову невизначеності. Перша пов'язана з нестачею інформації про об'єкти, явища і процеси, що відбуваються в природі, а друга - з поведінкою людини (або ОПР, або того, хто їй протидіє).

Поділ невизначеностей на природні і поведінкові носить штучний характер і враховує розходження в поведінці природи і людини при аналізі

можливих варіантів протидії з метою активного впливу на кінцевий результат.

Якщо дослідник прагне оцінити і урахувати вплив природних факторів на прийняття рішення для досягнення бажаного результату, то він виходить з припущення, що природа нейтральна стосовно дій людини, і найбільш розумним буде припущення про рівноможливість всіх розглянутих станів природи, тобто рівноможливості варіантів протидій при грі з природою.

При аналізі поведінкової невизначеності, тобто при грі з людиною, логічніше припустити, що своєю поведінку супротивник (як фактор) організує найгіршим для нас (найкращим для себе) способом; і це треба враховувати при виборі критерію прийняття рішення.

### **Контрольні запитання**

1. Чому необхідне використання математики в економіці?
2. Дайте визначення поняттю «модель».
3. Що таке моделювання?
4. Що таке математична модель?
5. Як будують математичну модель економічного явища або об'єкта?

Наведіть приклад побудови й уточнення моделі.

6. Перелічіть і поясніть основні принципи моделювання.
7. Який зв'язок між моделлю і ціллю системи?
8. У чому відмінність статичних моделей від динамічних?
9. Що таке прагматична модель? Наведіть декілька прикладів практичного застосування таких моделей.
10. Дайте визначення ідеальним моделям. Наведіть приклади таких моделей.
11. Перелічіть й охарактеризуйте основні властивості якості моделі.
12. Що таке «адекватна модель»?
13. Дайте визначення поняттям «рішення» і «прийняття рішення».
14. Сформулюйте послідовність процедур, які необхідно виконати для прийняття рішення. Чи можна змінити цю послідовність? Які з процедур можуть виконуватися паралельно?
15. Охарактеризуйте дві частини теорії прийняття рішень і перелічіть основні завдання, які вони вирішують.
16. Сформулюйте основні постулати теорії прийняття оптимальних рішень.
17. Дайте характеристику основним видам невизначеностей, що виникають у процесі прийняття рішень.
18. Охарактеризуйте основні напрямки психологічної теорії прийняття рішень.
19. Перелічіть основні методи прийняття рішень і сформулюйте ситуації, у яких ці методи можуть бути реалізовані.

## ТЕМА 2

# ПОНЯТТЯ ОПТИМІЗАЦІЙНИХ ЗАДАЧ І ОПТИМІЗАЦІЙНИХ МОДЕЛЕЙ. КЛАСИФІКАЦІЯ МЕТОДІВ

### 2.1 Основні поняття оптимізаційних задач і моделей

Економічні задачі, ціль яких полягає в знаходженні найкращого (оптимального) з погляду певного критерію або критеріїв варіанту використання наявних ресурсів (праці, капіталу та ін.), називають оптимізаційними.

Оптимізаційні задачі вирішують за допомогою оптимізаційних моделей за методами математичного програмування.

Структура оптимізаційної моделі складається з цільової функції, області припустимих рішень і системи обмежень, що визначають цю область. Цільова функція в самому загальному вигляді, у свою чергу, також складається з трьох елементів: керованих змінних, некерованих змінних, форми функції (виду залежності між ними).

Розглядаючи певну довільну систему рівнянь, яку складають  $m$  рівнянь з  $n$  невідомими, можна виділити три типи задач:

Якщо  $m = n$  - алгебраїчна задача, яка має одно рішення.

Якщо  $m > n$  - задача перевизначена і, як правило, не має рішення.

Якщо  $m < n$  - задача невизначена і має нескінченно багато рішень.

В економіці доводиться мати справу з задачами третього типу, тобто з такими, що мають нескінченно багато рішень.

Наведемо приклади оптимізаційних задач.

План постачань підприємств. Є ряд підприємств, що споживають відомі види сировини, і є ряд сировинних баз, що можуть поставляти цю сировину підприємствам. Бази зв'язані з підприємствами певними шляхами сполучення (залізничі, водний, автомобільний транспорт) із своїми тарифами. Потрібно розробити такий план постачань підприємств сировиною (з якої бази, в якій кількості, яку сировину доставити), щоб потреби в сировині були повністю забезпечені при мінімальних витратах на перевезення.

Будівництво ділянки магістралі. Споруджується ділянка залізничної магістралі. В нашому розпорядженні певна кількість ресурсів: людей, будівельних машин і механізмів, ремонтних майстерень, вантажних автомобілів тощо. Потрібно так спланувати будівництво, тобто призначити черговість робіт, розподілити машини і людей по ділянкам шляху, забезпечити ремонтні роботи, щоб воно було завершене в мінімально можливий термін при відповідній якості робіт.

Снігозахист залізниць. В деяких районах замети снігом залізничі складають серйозну заваду руху поїздів. Будь-яка перерва руху призводить до економічних втрат. Існує ряд можливих засобів снігозахисту (профіль дороги, захисні щити, застосування снігоочищувачів та ін.), кожний з яких вимагає відомих витрат на спорудження та експлуатацію. Потрібно розробити найбільш ефективні економічні засоби снігозахисту.

Бібліотечне обслуговування. Велика бібліотека обслуговує запити, що надходять від абонентів. В фондах бібліотеки є книги, що користуються підвищеним попитом, книги, на які запити надходять рідко і, нарешті, книги, на які запити майже ніколи не надходять. Є ряд можливостей розподілу книг на стелажах і сховищах, а також з диспетчеризації запитань із зверненнями до інших бібліотек. Потрібно розробити таку систему бібліотечного обслуговування, при якій запити абонентів задовольнялися б в максимальному ступені.

Оптимізаційну задачу можна вважати сформульованою математично, якщо сформульовано ціль управління, виражену через критерій ефективності та визначені обмеження, що представляють собою систему алгебраїчних рівнянь або нерівностей, які виражають обмеженість ресурсів або інших величин, використовуваних при управлінні.

Рішення про спосіб управління, що задовольняє всім поставленим обмеженням і перетворює на мінімум (максимум) критерій ефективності, називається **оптимальним рішенням**.

Отже, сутність всіх оптимізаційних задач збігається до пошуку такого розв'язку  $\bar{X}$ , що перетворює на екстремум критерій ефективності, виражений як функція від елементів прийнятого розв'язку  $\bar{X}$  й називаний **цільовою функцією**  $F(x)$ . Потрібно відзначити, що методи розв'язання оптимізаційних задач є не аналітичною, а алгоритмічною формою розв'язання задач, тобто дають не формулу, яка виражає остаточний результат, а вказують лише обчислювальну процедуру, що призводить до розв'язання задачі. Тому оптимізаційні методи стають ефективними головним чином при використанні обчислювальної техніки.

Відзначимо, що до оптимізаційних задач, як правило, незастосовні методи класичного аналізу для відшукування умовних екстремумів. Це пов'язано з такими специфічними їхніми особливостями:

- 1) коли на елементи розв'язку  $\bar{X}$  накладені обмеження, екстремум часто досягається не в точках, де похідні дорівнюють нулю, а на границі області обмежень;
- 2) у практичних задачах число змінних і число обмежень настільки велике, що пошук екстремуму шляхом визначення похідних стає не ефективним;
- 3) у багатьох задачах математичного програмування цільова функція взагалі не має похідних (наприклад, задана тільки для цілочислових значень аргументів).

У зв'язку з цим метою математичного програмування є створення аналітичних методів визначення розв'язку або ефективних обчислювальних способів одержання наближеного розв'язку оптимізаційної задачі.

Математичне моделювання економічних процесів є, з одного боку, дуже важливим і складним, а з іншого боку – таким, що практично не піддається науковій формалізації. Неодноразові спроби, виділити загальні принципи створення математичних моделей призводили або до декларування рекомендацій самого загального характеру, важко застосовних для розв'язання конкретних проблем, або, навпаки, до появи рецептів, застосовних у дійсності



тільки до вузького кола задач. Тому більш корисним представляється знайомство з технікою математичного моделювання на конкретних прикладах.

Спільним для оптимізаційних задач є те, що в них стоїть проблема пошуку найбільшого або найменшого (**оптимального**) значення певної функції, що відбиває **ціль управління** системою, або, як ще кажуть, цільової функції. Пошук оптимального значення здійснюється на певній підмножині припустимих значень змінних, що описують стан цієї системи, іменованій **множиною припустимих планів**.

Нехай на певній множині  $D$  визначена функція  $f(x)$ . Нагадаємо, що точка  $x^*$ , що належить  $D$  ( $x^* \in D$ ), називається **точкою глобального максимуму**, якщо для будь-якого  $x \in D$  виконується нерівність  $f(x) \leq f(x^*)$ . У цьому випадку значення  $f(x^*)$  називається **глобальним максимумом функції**. Точка  $\hat{x}$  називається **точкою локального максимуму**, якщо існує певне оточення цієї точки, у будь-якій точці якого значення функції менші, ніж в  $\hat{x}$  ( $f(x) \leq f(\hat{x})$ ). Аналогічно визначаються **глобальний і локальний мінімуми**. Узагальнюючим поняттям для максимуму й мінімуму є такий термін, як **екстремум (оптимум)**.

Необхідно відзначити, що далеко не завжди весь комплекс цілей і задач, що стоїть перед об'єктом, який моделюється, може бути виражений у формі певної цільової функції. Більше того, усвідомлення й осмислення цієї проблеми стало свого роду переломним етапом в історії розвитку даної науки, що дали поштовх до розвитку нових напрямків, пов'язаних з методами **багатоокритеріальної (або векторної) оптимізації**, коли критерієм оптимальності є вимога мінімізації або максимізації кількох скалярних функцій. Проте всі вони базуються на методах **однокритеріальної оптимізації**, без ясного розуміння яких неможлива робота із складнішим математичним апаратом.

Потужним інструментом вирішення подібного роду задач стали спеціальні методи пошуку екстремуму, що складають зміст математичного програмування. У цьому випадку поняття програмування вживається в значенні планування (на відміну від програмування для ЕОМ).

Задачі математичного програмування мають велику розмаїтість. Математичне моделювання таких задач практично не піддається науковій формалізації через те, що принцип побудови математичної моделі істотно залежить від конкретної природи досліджуваної системи. Проте у цих задачах прийнято виділяти певну послідовність етапів дослідження:

1. Постановка задачі.
2. Словесне формулювання задачі з визначенням цілі її розв'язання та факторів-обмежень, що впливають на нього (вербальна модель);
3. Формалізація задачі – побудова адекватної математичної моделі. На цьому етапі цільова функція  $F(x)$  виражається як залежність від розв'язку  $\bar{X}$ , а обмеження записують у вигляді системи рівностей і нерівностей;
4. Розв'язання задачі на базі математичної моделі;

5. Перевірка отриманих результатів на їхню адекватність природі досліджуваної системи, можливе коректування первісної моделі.
6. Розробка рекомендацій на підставі отриманого розв'язку.

Введемо ряд визначень.

**Розв'язком** (або **планом**) називається усякий певний вибір параметрів, що залежать від нас. Параметри, сукупність яких утворює розв'язок, називаються **елементами розв'язку**. Як елементи розв'язку можуть фігурувати числа, вектори, функції, фізичні ознаки та ін.

$$x = (x_1, x_2, \dots, x_j, \dots, x_n).$$

Система обмежень за ресурсами формує множину припустимих розв'язків (планів)  $D$ . Той факт, що розв'язок  $x$  належить множині припустимих розв'язків  $D$ , записується в такий спосіб:

$$x \in D.$$

**Оптимальним розв'язком** або **оптимальним планом** називають таке рішення, що обертає цільову функцію  $F(x)$  на максимум або мінімум.

Отже, у найбільш загальному вигляді оптимізаційна задача формулюється в такий спосіб:

При заданих обмеженнях знайти таке рішення  $x = x^*$ , що обертає цільову функцію  $F(x)$  на максимум або мінімум.

$$F^* = \underset{x \in D}{extr} \{F(x, \alpha)\},$$

де  $\alpha$  - система обмежень задачі.

## 2.2 Класифікація методів оптимізації

Одним з найбільш загальних методів розв'язання задач пошуку екстремуму функції при наявності обмежень на її змінні (задачі умовної оптимізації) є метод Лагранжа, відомий з курсу диференціального обчислення, ідея якого полягає у зведенні задачі пошуку умовного екстремуму цільової функції до задачі безумовної оптимізації.

Певна специфіка економічних оптимізаційних задач з обмеженнями, призвела до перегляду класичних методів і створенню нових методів, відомих під назвою методів програмування.

Залежно від вигляду цільової функції  $F(x)$  і системи обмежень  $\alpha$  виділяють наступні методи розв'язання оптимізаційних задач.

**Лінійне програмування.** Застосовується, якщо в моделі цільова функція  $F(x)$  є лінійною, а множина  $D$ , на якій шукається її екстремум, задається системою лінійних рівнянь і нерівностей. У лінійному програмуванні існує клас задач, структура яких дозволяє створити спеціальні методи розв'язання, що вигідно відрізняються від методів розв'язання задач загального характеру. Зокрема - транспортна задача.

**Нелінійне програмування.** Тут є нелінійними цільова функція й обмеження. У нелінійному програмуванні виділяють такі класи задач:

- **опукле програмування** – коли опукла цільова функція (якщо розглядається задача її мінімізації) і опукла множина, на якій вирішується екстремальна задача;

- **квадратичне програмування** – коли цільова функція є квадратичною, а обмеження – лінійні рівності або нерівності.

**Дискретне програмування.** Даний метод використовують, коли на елементи рішення  $x$  накладена вимога дискретності, наприклад, цілочисловості. Така вимога істотно ускладнює розв'язання задачі, тому що застосування стандартних прийомів (вирішити задачу як аналогову, а потім округлити результат до цілого значення) неможливо.

**Динамічне програмування.** Це метод, що дозволяє шляхом покрокової оптимізації деяких проміжних цільових функцій отримати загальний результуючий оптимум. У задачах динамічного програмування цільова функція  $F(x)$  є аддитивною або мультиплікативною функцією змінних  $x$ .

**Стохастичне програмування.** Даний вид програмування використовується, коли параметри умов або елементи розв'язку є випадковими величинами, що обумовлено невизначеністю, яка породжує ризикованість прийнятих рішень. У стохастичному програмуванні труднощі виникають не тільки при розробці методів розв'язання задач, а й при їхній постановці.

**Евристичне програмування.** Застосовують для розв'язання задач, у яких точний оптимум знайти алгоритмічним шляхом неможливо через величезне число варіантів. У такому випадку відшукують не оптимальний, а досить гарний з погляду практики розв'язок.

## Контрольні запитання

1. Охарактеризуйте особливості оптимізаційних задач.
2. Які загальні етапи розв'язання оптимізаційних задач прийнято виділяти?
3. Чому до оптимізаційних задач не застосовують класичні методи пошуку умовного екстремуму функції?
4. Що являє собою цільова функція оптимізаційної задачі? Яке її призначення?
5. Дайте визначення понять: план, припустимий план, оптимальний план, розв'язок оптимізаційної задачі.
6. На чому заснована класифікація моделей і методів математичного програмування з розв'язання оптимізаційних задач? Які класи моделей і методів виділяють у математичному програмуванні?
7. Що являє собою множина можливих розв'язків задачі математичного програмування?
8. Поясніть, яка область можливих розв'язків задачі математичного програмування називається областю припустимих планів.

### ТЕМА 3 ЛІНІЙНЕ ПРОГРАМУВАННЯ

Серед задач математичного програмування найпростішими й найкраще розробленими є задачі лінійного програмування. Характерним для них є те, що:

- цільова функція  $F(x)$  лінійно залежить від елементів розв'язку  $x_1, x_2, \dots, x_n$ . Її називають **лінійною формою** й позначають  $L$ ;
- обмеження, що накладають на елементи розв'язку, мають вигляд лінійних рівностей і нерівностей відносно  $x_1, x_2, \dots, \dots, x_j, \dots, x_n$ .

Такі задачі часто зустрічаються на практиці. До задач лінійного програмування належить, зокрема, розглянута раніше найпростіша задача виробничого планування (задача про оптимальне використання ресурсів). Також до задач лінійного програмування належать задачі про використання інвестицій, про мінімізацію витрат, транспортна задача, задача про складання раціону та ін.

Математичну модель задачі лінійного програмування завжди записують у двох формах – у **загальній формі** (ЗЗЛП) і **канонічній формі** (КЗЛП).

#### 3.1 Загальна форма задачі лінійного програмування (ЗЗЛП)

У загальному вигляді задача лінійного програмування (ЗЛП) формулюється в такий спосіб:

Знайти найбільше або найменше значення лінійної функції

$$L = \sum_{j=1}^n c_j x_j \rightarrow \text{extr} . \quad (3.1)$$

на деякій множині  $D$ , де  $x \in D$  задовольняє системі обмежень

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\leq b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\leq b_2, \\ &\dots\dots\dots \\ a_{k1}x_1 + a_{k2}x_2 + \dots + a_{kn}x_n &\leq b_k, \\ a_{(k+1)1}x_1 + a_{(k+1)2}x_2 + \dots + a_{(k+1)n}x_n &= b_{(k+1)}, \\ &\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m, \\ x_j &\geq 0, \quad j = \overline{1, n} . \end{aligned} \quad (3.2)$$

Відзначимо, що в системі перші  $k$  обмежень є нерівностями, а наступні  $m-k$  – рівняннями. Цього можна домогтися простим переупорядкуванням виразів.

Щодо направлення знака нерівності будемо вважати, що ліва частина менша або дорівнює правій частині. Домогтися цього можна, помноживши на  $(-1)$  обидві частини нерівностей із протилежним знаком.

Вибір типу шуканого екстремуму цільової функції також не принциповий, оскільки задача пошуку максимуму функції  $L = \sum c_j x_j$  є еквівалентною задачі пошуку мінімуму функції  $-L = \sum (-c_j x_j)$ .

Задачу лінійного програмування, записану в такій формі, називають загальною задачею лінійного програмування (ЗЛП).

### 3.2 Основні властивості ЗЛП і її перша геометрична інтерпретація

**Основні поняття лінійної алгебри та опуклого аналізу, застосовувані в теорії математичного програмування.** Коротко нагадаємо деякі фундаментальні визначення й теореми лінійної алгебри й опуклого аналізу, які широко застосовуються при розв'язанні задач як лінійного, так і нелінійного програмування.

Фундаментальним поняттям лінійної алгебри є лінійний (речовинний) простір. Під ним мають на увазі множину деяких елементів (іменованих векторами або точками), для яких задані операції додавання й множення на речовинне число (скаляр), причому елементи, що є результатом виконання операцій, також відповідно до визначення повинні належати вихідному простору.

Окремими випадками лінійних просторів є речовинна пряма, площина, геометричний тривимірний простір.

Вектор  $\lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_m a_m$  називається **лінійною комбінацією** векторів  $a_1, a_2, \dots, a_m$  з коефіцієнтами  $\lambda_1, \lambda_2, \dots, \lambda_m$ .

Система векторів лінійного простору  $a_1, a_2, \dots, a_m$  називається **лінійно залежною**, якщо існують такі числа  $\lambda_1, \lambda_2, \dots, \lambda_m$ , які не дорівнюють одночасно нулю, що їхня лінійна комбінація  $\lambda_1 a_1, \lambda_2 a_2, \dots, \lambda_m a_m$  дорівнює нульовому вектору (вектору, усі компоненти якого дорівнюють нулю). У протилежному випадку систему  $a_1, a_2, \dots, a_m$  називають **лінійно незалежною**, тобто лінійна комбінація даних векторів може дорівнювати нульовому вектору тільки при нульових коефіцієнтах  $\lambda_1, \lambda_2, \dots, \lambda_m$ .

Максимально можлива кількість векторів, які можуть утворювати лінійно незалежну систему в даному лінійному просторі, називають **розмірністю простору**, а будь-яку систему лінійно незалежних векторів у кількості, рівній розмірності, - **базисом простору**.

Лінійний простір позначають  $R^n$ , де  $n$  - його розмірність.

Будь-яка підмножина даного лінійного простору, що саме має властивості лінійного простору, називається лінійним підпростором. Множина  $H$ , одержувана зсувом деякого лінійного підпростору  $L \in R^n$  на вектор  $a \in R^n$ :  $H = L + a$ , називається афінною множиною (простором). Якщо фундаментальною властивістю будь-якого лінійного простору або підпростору є приналежність йому нульового вектора, то для афінної множини це не завжди так. На площині прикладом підпростору є пряма, яка проходить через початок координат, а афінної множини - будь-яка пряма на площині. Характерною властивістю афінної множини є належність їй будь-якої прямої, що з'єднує дві будь-які її точки. Розмірність афінної множини збігається з розмірністю того лінійного підпростору, зсувом якого її отримано.

Якщо розглядається деякий лінійний простір  $R^n$ , то приналежні йому афінні множини розмірності 1 називаються прямими, а розмірності  $(n - 1)$  — гіперплощинами. Так, звичайна площина є гіперплощиною для тривимірного геометричного простору  $R^3$ , а пряма — гіперплощиною для площини  $R^2$ . Усяка гіперплощина поділяє лінійний простір на два півпростори.

Множина  $V$  векторів (точок) лінійного простору  $R^n$  називається **опуклою**, якщо вона містить відрізок прямої, яка з'єднує дві його будь-які точки, або, інакше кажучи, з того, що  $a \in V$  і  $b \in V$ , випливає, що  $x = (1 - \lambda)a + \lambda b \in V$ , де  $0 \leq \lambda \leq 1$ .

Лінійна комбінація  $\sum_{i=1}^m \lambda_i a_i$  векторів  $a_1, a_2, \dots, a_m$  називається **опуклою**, якщо  $\lambda_i \geq 0$ ,  $i = \overline{1, m}$  і  $\sum_{i=1}^m \lambda_i = 1$ .

Множину, що містить всі можливі опуклі комбінації точок деякої множини  $M$ , називають опуклою оболонкою даної множини. Можна показати, що опукла оболонка множини  $M$  є найменшою опуклою множиною, що містить  $M$ .

Опукла оболонка кінцевої множини точок називається **опуклим багатогранником**, а непусте перетинання кінцевого числа замкнутих півпросторів — **багатогранною опуклою множиною**. На відміну від опуклого багатогранника останнє може бути необмеженим.

Точка  $v$  опуклої множини  $V$  називається його кутовою (крайньою) точкою, якщо вона не є внутрішньою точкою ні для якого відрізка, кінці якого належать множині  $V$ . Кутові точки опуклого багатогранника є його вершинами, а сам він — опуклою оболонкою своїх вершин.

**Перша геометрична інтерпретація ЗЛП і графічний метод розв'язання.** У тому випадку, коли ЗЛП містить дві змінні  $x_1$  і  $x_2$ , її можна зобразити на координатній площині й одержати розв'язок графічним методом. Графічне розв'язання ЗЛП носить ілюстративний характер, але основний зміст і термінологія розповсюджуються на задачі великої розмірності.

Розглянемо приклад. Нехай цільова функція представлена виразом

$$L = x_1 + 3x_2 \rightarrow \max,$$

а обмеження задані системою нерівностей:

$$\begin{aligned} x_1 + x_2 &\leq 6, \\ x_1 - x_2 &\leq 2, \\ x_1 &\leq 3, \\ x_1 &\geq 0, x_2 \geq 0. \end{aligned}$$

Зобразимо пару значень  $x_1$  та  $x_2$  точкою на координатній площині  $x_1 O x_2$  з координатами  $(x_1, x_2)$ , що показано на рисунку 3.1.

Кожна нерівність визначає певну напівплощину. Перетинання трьох напівплощин являє собою множину припустимих планів  $D$ , тому що кожна точка його множини належить одночасно кожній з трьох напівплощин, а отже задовольняє обмеженням ЗЛП. Помітимо, що припустимих розв'язків — нескінченна множина.

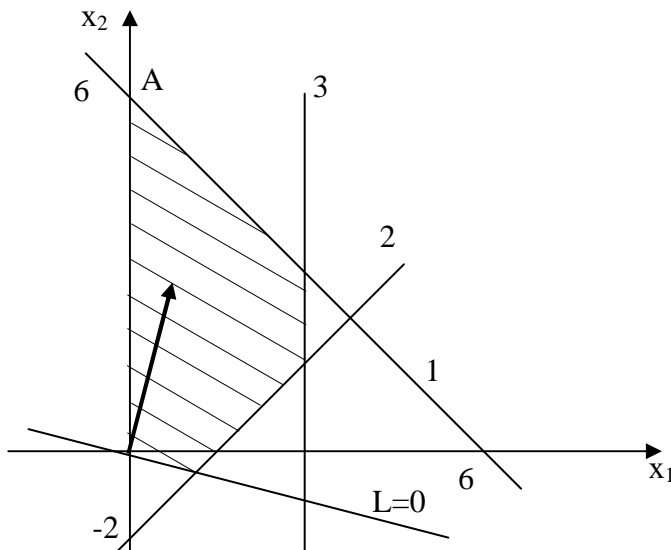


Рисунок 3.1 – Геометрична інтерпретація ЗЗЛП

Для визначення оптимального плану задачі, тобто такого розв'язку  $(x_1, x_2)$ , що обертає цільову функцію на максимум, скористаємося наступними визначеннями:

- градієнтом функції  $f(x)$  називається вектор

$$\nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right),$$

який вказує напрямок найбільш швидкого зростання функції  $f(x)$ ;

- лінією рівня функції  $f(x)$  називається множина точок з області її визначення, у яких функція приймає те саме фіксоване значення.

У нашому прикладі  $\nabla L = (1, 3)$ . Лінії рівня  $L$  перпендикулярні напрямку градієнта. Побудуємо опорну пряму  $L=0$ , що проходить через початок координат, і будемо переміщати її в напрямку  $\nabla L$ . Очевидно, що переміщати лінію рівня в напрямку зростання цільової функції має сенс тільки в межах області припустимих розв'язків. Точкою, у якій цільова функція дістане максимального значення, у нашому прикладі є точка  $A$  з координатами  $(0, 6)$ . Отже, отриманий оптимальний план задачі має вигляд

$$x^* = (0, 6),$$

він зумовлює максимальне значення цільової функції

$$L = 1 \cdot 0 + 3 \cdot 6 = 18.$$

Теоретично можливі також наступні окремі випадки розв'язку ЗЗЛП:

- цільова функція  $L$  не обмежена зверху, тобто не має максимуму (рис. 3.2);
- коли лінія рівня збігається із гранню області припустимих розв'язків (рис. 3.3). У цьому випадку всі точки, що лежать на грані множини  $D$ , є оптимальними планами й кажуть, що має місце **альтернативний оптимум**.

У розглянутих ілюстраціях припустимі плани ЗЗЛП мають вигляд опуклої багатогранної множини. Таке подання множини припустимих планів називається першою геометричною інтерпретацією ЗЗЛП.

**Основні теореми лінійного програмування.** Розглянемо деякі теореми, що відбивають фундаментальні властивості задач лінійного програмування і полягають в основі методів їхнього розв'язання. Вони узагальнюють на випадок задач із довільною кількістю змінних ті властивості, які ми спостерігали у двовимірному випадку.

**Теорема 3.1.** Якщо цільова функція  $L$  приймає максимальне значення в певній точці множини припустимих планів  $D$ , то вона приймає це значення й у певній кутовій точці даної множини.

Доказ.

Щоб не ускладнювати виклад, обмежимося тим випадком, коли множина  $D$  є обмеженою, і, отже, є опуклим багатогранником.

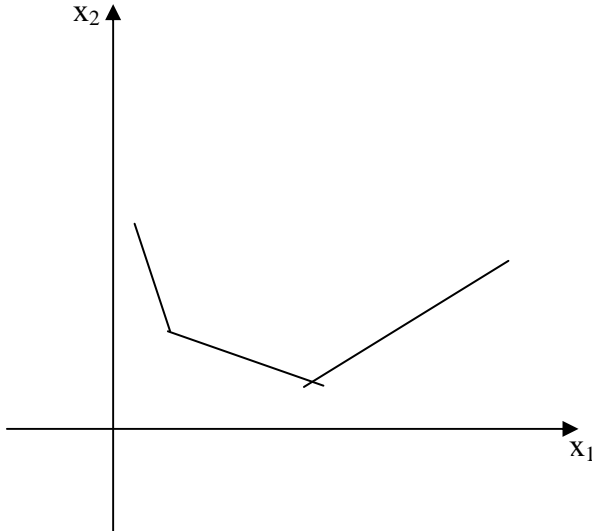


Рисунок 3.2 – Цільова функція  $L$  не обмежена зверху

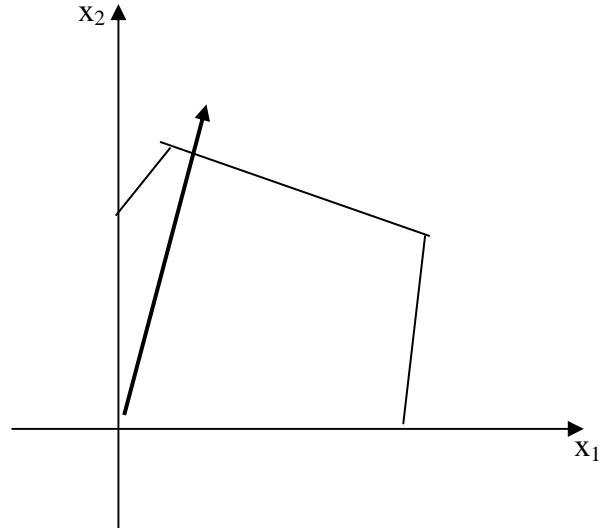


Рисунок 3.3 – Альтернативний оптимум

Для доказу скористаємося наступною відомою властивістю обмежених опуклих множин: якщо  $D$  - замкнена обмежена опукла множина, що має кінцеве число кутових точок, то будь-яку точку  $x \in D$  можна подати у вигляді опуклої комбінації кутових точок  $D$ .

Нехай  $x_1, x_2, \dots, x_m$  - кутові точки множини  $D$ , а  $x^*$  - точка, у якій цільова функція  $L$  досягає максимуму. Точку  $x^*$  можна подати у вигляді опуклої комбінації кутових точок  $x_1, x_2, \dots, x_m$

$$x^* = \sum_{i=1}^m \lambda_i x_i,$$

де  $\sum \lambda_i = 1$ ,  $\lambda_i \geq 0$ ,  $i = \overline{1, m}$ .

Оскільки  $x^*$  - точка максимуму, то для будь-якого  $x$   $cx^* \geq cx$ . У тому числі й для  $cx_r$  ( $x_r$  - кутова точка).

Функція  $L(x)$  - лінійна, тому

$$L\left(\sum_{i=1}^m \lambda_i x_i\right) = \sum_{i=1}^m \lambda_i L(x_i),$$

а значить

$$cx^* = c \sum_{i=1}^m \lambda_i x_i = \sum_{i=1}^m \lambda_i (cx_i) \leq \sum_{i=1}^m \lambda_i (cx_r) = cx_r \sum_{i=1}^m \lambda_i = cx_r,$$

де  $x_r$  - кутова точка множини  $D$ , що задовольняє умові

$$cx_r = \max_{1 \leq i \leq m} \{cx_i\}.$$



Таким чином,  $cx^* \leq cx_r$ . У той же час  $cx^* \geq cx_r$ , звідки випливає  $cx^* = cx_r$ .

Тобто існує принаймні одна кутова точка  $x_r$ , у якій цільова функція приймає максимальне значення.

**Теорема 3.2.** Якщо цільова функція  $L$  приймає максимальне значення в декількох точках множини  $D$ , то вона приймає це ж значення в будь-якій точці, що є їхньою опуклою комбінацією.

Доказ.

Нехай максимальне значення цільової функції  $L(x)$  досягається в точках  $x_1, x_2, \dots, x_s$ , тобто  $L^* = cx_i$ ,  $i = \overline{1, s}$ . Розглянемо довільну опуклу комбінацію цих точок

$$x^* = \sum_{i=1}^s \lambda_i x_i,$$

де  $\sum \lambda_i = 1$ ,  $\lambda_i \geq 0$ ,  $i = \overline{1, s}$ .

Знайдемо значення цільової функції в точці  $x^*$

$$L(x^*) = cx^* = c \sum_{i=1}^s \lambda_i x_i = \sum \lambda_i (cx_i) = \sum \lambda_i L^* = L^* \sum \lambda_i = L^*.$$

Теорема 3.1 є фундаментальною, тому що вона вказує принциповий шлях розв'язання ЗЛП. Замість дослідження нескінченної множини припустимих розв'язків для знаходження серед них оптимального, необхідно досліджувати лише кінцеве число кутових точок багатогранника розв'язків.

### 3.3 Канонічна форма задачі лінійного програмування (КЗЛП)

Канонічною називається задача лінійного програмування, якщо всі її обмеження є рівняннями.

Переважає більшість методів розв'язання задач лінійного програмування призначена для канонічних задач. Тому початковий етап розв'язання будь-якої загальної ЗЛП завжди пов'язаний із приведенням її до еквівалентної КЗЛП.

Загальна ідея переходу від ЗЛП до КЗЛП досить проста. Обмеження у вигляді нерівностей перетворюють на рівняння за рахунок додавання фіктивних невід'ємних змінних, які одночасно входять до цільової функції з коефіцієнтом 0, тобто не надають впливу на її значення

$$x'_{n+i}, \quad i = \overline{1, k}.$$

Змінні, на які не накладено умову невід'ємності, подаються у вигляді різниці двох нових невід'ємних змінних.

Додатково треба помітити, що вибір типу шуканого екстремуму (максимуму або мінімуму) носить відносний характер. Так, задача пошуку максимуму функції

$$L(x) = \sum_{j=1}^n c_j x_j$$

еквівалентна задачі пошуку мінімуму функції

$$-L(x) = \sum_{j=1}^n -c_j x_j.$$

### Знайти найбільше або найменше значення функції

на деякій множині  $D$ , де  $x \in D$  задовольняє системі обмежень

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + x_{n+2} = b_2,$$

$$a_{k1}x_1 + a_{k2}x_2 + \dots + a_{kn}x_n + x'_{n+k} = b_k, \quad (3.4)$$

$$a_{k+1,1}x_1 + a_{k+1,2}x_2 + \dots + a_{k+1,n}x_n = b_{k+1},$$

.....

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m,$$

$$x_j \geq 0, \quad j = \overline{1, n+k}.$$

Таким чином, «платою» за перехід від загальної форми задачі лінійного програмування до канонічної є зростання її розмірності, що, за інших рівних умов, ускладнює процес розв'язання.

Запишемо КЗЛП у матричній формі:

$$\overline{C}^T \overline{X} \quad (3.5)$$
$$\overline{A}^T \overline{X} = \overline{B}, \quad \overline{X} \geq 0, \quad (3.6)$$
$$\overline{A}_I = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}; \quad \overline{A}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}; \quad \overline{A}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}; \quad \overline{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}. \quad (3.7)$$

Вектори  $\overline{A}_j$ ,  $j = \overline{1, n}$  називають векторами вимог задачі. Вектор  $\overline{B}$  називається вектором обмежень. При цьому компоненти деякого припустимого вектора  $x \in D$  —  $x_j$  — є коефіцієнтами розкладання вектора обмежень  $\overline{B}$  задачі по векторах вимог  $\overline{A}_j$ .

$$\overline{A} \cdot \overline{X} = \overline{B}$$

26

Якщо деякі  $m$  стовпців  $A_1, A_2, \dots, A_m$  матриці  $\bar{A}$  є лінійно незалежними, то вони утворюють **базис** у просторі  $R^m$  і їх буде достатньо для подання вектора  $\bar{B}$  у вигляді лінійної комбінації зазначених стовпців. Це означає, що інші стовпці ввійдуть до даного розкладання з нульовими коефіцієнтами.

Якщо коефіцієнти лінійної комбінації виявляються невід'ємними, то ми отримаємо **базисний припустимий план**  $x$ , у якого не більше за  $m$  елементів відмінні від нуля. Або іншими словами. Система обмежень КЗЛП являє собою систему  $m$  рівнянь з  $n$  змінними, причому  $m \leq n$ .

У такій системі  $m$  змінних називаються **базисними**, а інші  $(n-m)$  змінних – **вільними**.

**Базисним розв'язком** системи  $m$  лінійних рівнянь з  $n$  змінними називається розв'язок, у якому всі  $n-m$  вільних змінних дорівнюють нулю.

Базисними розв'язками можуть бути різні групи з  $n$  змінних. У загальному випадку число таких груп не перевершує  $C_n^m$ .

Таким чином, система з  $m$  лінійних рівнянь з  $n$  змінними ( $m < n$ ) є **невизначеною**, тому що кожному довільному набору вільних змінних відповідає один базисний розв'язок системи.

**Опорним планом** КЗЛП називається припустимий базисний розв'язок, компоненти якого більше нуля.

Базисний розв'язок, у якому хоча б одна з базисних змінних дорівнює нулю, називається **виродженням**.

### Властивості базисних планів задачі лінійного програмування.

**Теорема 3.3.** Кожний припустимий базисний план є кутовою точкою множини припустимих планів  $D$ .

Доказ:

Будемо вважати, що базисними є перші  $m$  стовпців матриці  $\bar{A}$   
 $A_1, A_2, \dots, A_m \dots$

Тоді формулювання теореми можна записати у вигляді:

Якщо існує такий  $n$ -мірний вектор

$$x = \left( x_1, x_2, \dots, x_k, \underbrace{0, \dots, 0}_{n-k} \right),$$

$$x_j > 0; \quad j = \overline{1, k}; \quad k \leq m,$$

що  $x_1 A_1 + x_2 A_2 + \dots + x_k A_k = B_0$ , то  $x$  є кутовою точкою множини  $D$ .

Припустимо, що базисний план  $x$  не є кутовою точкою множини  $D$ . У цьому випадку його можна уявити у вигляді опуклої комбінації двох різних припустимих планів  $x^1$  і  $x^2$

$$x = \lambda x^1 + (1 - \lambda) x^2, \quad 0 < \lambda < 1.$$

Оскільки останні  $(n-k)$  компонент вектора  $x$  дорівнюють нулю, а  $\lambda$  і  $(1-\lambda)$  більше нуля, то ці ж  $(n-k)$  компонент у векторах  $x^1$  і  $x^2$  також дорівнюють нулю.

Оскільки плани  $x^1$  і  $x^2$  – припустимі плани, маємо

$$x_1^1 A_1 + x_2^1 A_2 + \dots + x_k^1 A_k = B,$$

$$x_1^2 A_1 + x_2^2 A_2 + \dots + x_k^2 A_k = B.$$

Відніmemo з першого рівняння друге

$$(x_1^1 - x_1^2)A_1 + (x_2^1 - x_2^2)A_2 + \dots + (x_k^1 - x_k^2)A_k = 0.$$

Отримали нульовий вектор. Оскільки  $A_j$  лінійно незалежні, нулю дорівнюють коефіцієнти

$$(x_1^1 - x_1^2) = 0$$

$$(x_2^1 - x_2^2) = 0$$

$$\vdots$$

$$(x_k^1 - x_k^2) = 0$$

Звідки випливає, що  $x^1 = x^2$ . Це суперечить припущенню, що  $x^1$  і  $x^2$  є різними кутовими точками множини  $D$ .

Таким чином, план  $x$  не може бути поданим у вигляді опуклої комбінації двох інших точок множини, отже є кутовою точкою даної множини.

Справедливо й зворотнє ствердження: Якщо  $x$  - кутова точка множини  $D$ , то вона є припустимим базисним планом задачі лінійного програмування.

### 3.4 Симплекс-метод

На підставі розглянутих теорем про властивості лінійних екстремальних задач можна побачити, що пошук їхніх розв'язків зводиться до послідовного перебору кутових точок множини припустимих планів.

Однак такий перебір для реальних багатомірних задач на практиці нездійсненний або вкрай неефективний. Наприклад, число базисних планів у задачі з 10 змінними й 30 обмеженнями становить

$$C_{30}^{10} = \frac{30!}{10! \cdot 20!} = 1489411.$$

Класичним методом розв'язання задач лінійного програмування є симплекс-метод, що також називається методом послідовного поліпшення плану. Він розроблений в 1947 році американським математиком Джорджем Данцигом.

Симплекс-метод являє собою послідовний перебір кутових точок області припустимих розв'язків, при якому значення цільової функції зростає від ітерації до ітерації (від однієї кутової точки до іншої).

Критерій оптимальності в симплекс-методі реалізується шляхом визначення спеціальних оцінок  $\Delta_j$  для небазисних векторів-стовпців матриці  $\bar{A}$ , щодо поточного базису (симплекс-різниць). Симплекс-різниці обчислюють за формулою

$$\Delta_j = L_j - c_j, \quad (3.8)$$

де  $L_j$  – індекси векторів, що відповідають поточному базису

$$L_j = \sum_{i=1}^m c_j a_{ij}. \quad (3.9)$$

Сформулюємо критерій оптимальності припустимого базисного плану:

**план є оптимальним, якщо для всіх  $j = \overline{1, n}$   $\Delta_j \geq 0$ , і неоптимальним у протилежному випадку, тобто якщо існує таке  $j = \overline{1, n}$ , що  $\Delta_j < 0$ .**

Якщо симплекс-різниці показують неоптимальність плану, здійснюється перехід до наступного базису. При цьому один стовпець виводиться з базису, а інший вводиться. Для забезпечення покращення значення цільової функції в базис повинен бути введеним вектор-стовпець, що має від'ємну оцінку. Якщо таких стовпців декілька, то для уведення рекомендується вибирати стовпець, що має максимальний за модулем добуток оцінки  $\Delta_j$  на відношення  $\Theta_r = \min_i \left\{ \frac{b_i}{a_{ij}} \right\}$ . Одночасно на цьому кроці потрібно ухвалити

рішення щодо того, який стовпець треба вивести з базису. Зробити це потрібно таким чином, щоб знову формований базис опинився припустимим.

Можна довести, що допустимість базису, до якого здійснюється перехід, забезпечується наступним правилом виводу стовпця з поточного базису:

для стовпця, що претендує на введення до базису, і вектора обмежень  $\bar{B}$  розглядаються відношення

$$\Theta_i = \frac{b_i}{a_{ij}} \quad (3.10)$$

і визначають такий рядок  $r$ , що

$$\Theta_r = \min_i \{\Theta_i\}. \quad (3.11)$$

Отриманий індекс  $r$  визначає номер рядка, який відповідає вектору, виведеному з базису.

Таким чином, якщо базис на  $q$ -ї ітерації включав стовпці з номерами

$$\{j_1, j_2, \dots, j_{r-1}, j_r, j_{r+1}, \dots, j_m\},$$

то базис на ітерації  $q + 1$  буде складатися зі стовпців з номерами:

$$\{j_1, j_2, \dots, j_{r-1}, j_{r+1}, \dots, j_m\}.$$

Окремо слід домовитися про випадок, коли стовпець, що претендує на введення до базису, не містить додатних компонентів ( $a_{ij} < 0$ ). Це означає, що цільова функція у задачі не обмежена на множині припустимих значень, тобто може досягати як завгодно великого значення, отже оптимальний план відсутній.

Після переходу до нового базису можна заново сформувати матрицю  $\bar{A}$  й дослідити новий план на оптимальність. З погляду техніки обчислень представляється раціональним безпосередньо переходити від матриці  $q$ -ї ітерації до матриці  $(q+1)$ -ї ітерації. Справа в тому, що в цих матрицях стовпці, які відповідають базисним векторам, складаються з нулів, за винятком одного елемента, що дорівнює одиниці. Позиція цього ненульового (одичного) елемента визначається порядковим номером базисного стовпця. Тому для одержання матриці  $(q+1)$ -ї ітерації досить за допомогою лінійних операцій над рядками матриці  $q$ -ї (попередньої) ітерації привести її стовпець, що відповідає вектору, який вводиться до базису, до «базисного» вигляду.

Для цього застосовується перетворення Жордана-Гауса (так званий метод повного виключення). У цьому випадку воно полягає в тому, що ми повинні отримати одиницю на місці елемента  $a_{rj}$  (він називається ведучим) і нулі на місці інших елементів стовпця  $a_{ij}$ . Перше досягається за допомогою поділення  $r$ -го рядка на ведучий елемент, друге – шляхом додавання знову отриманого  $r$ -го рядка, помноженого на відповідний коефіцієнт, до інших рядків матриці  $q$ -ї (попередньої) ітерації.

Формально результат виконання даного перетворення над елементами матриці може бути виражений у наступному виді:

$$a_{rj}^{q+1} = \frac{a_{rj}^q}{a_{rk}^q}, \quad b_r^{q+1} = \frac{b_r^q}{a_{rk}^q}, \quad (3.12)$$

де  $j = \overline{1, n}$ ;

$k$  – номер стовпця, що вводиться до базису;

$$a_{ij}^{q+1} = a_{ij}^q - a_{ik}^q \frac{a_{rj}^q}{a_{rk}^q}, \quad (3.13)$$

де  $i = \overline{1, m}$ ,  $i \neq r$ ;

$$b_i^{q+1} = b_i^q - a_{ik}^q \frac{b_r^q}{a_{rk}^q}, \quad (3.14)$$

де  $i = \overline{1, m}$ ,  $i \neq r$ .

Слід особливо зазначити зміст елементів вектора  $\overline{B}$ . Його нульовий компонент  $b_0$  містить значення цільової функції, що досягається нею на поточному плані, а інші елементи є ненульовими компонентами цього плану.

Приведемо схему алгоритму симплекс-методу для розв'язання задачі максимізації.

1. Знаходження припустимого базисного плану.

2. Перевірка оптимальності поточного базисного плану: здійснюється перегляд рядка оцінок  $\Delta_j$ . Можливі два варіанти:

-  $\Delta_j \geq 0$  - план, що відповідає поточному базису задачі, є оптимальним. Обчислювальний процес закінчений. Випишується оптимальний план задачі  $x^*$  і значення цільової функції  $L(x^*)$ ;

- у рядку оцінок  $\Delta_j$  існує щонайменше один елемент  $\Delta_j < 0$ , тобто оцінка якого є від'ємною. Отже, план є неоптимальним. Вибирається стовпець із номером  $k$ , для якого добуток  $\Delta_j \Theta$  найбільший за абсолютною величиною. Він називається ведучим і повинен бути введений до чергового базису;

3. Визначення стовпця, виведеного з базису. Досліджується ведучий стовпець, можливі два варіанти:

- для всіх  $i = \overline{1, m}$   $a_{ik}^q < 0$ . Робиться висновок про необмеженість цільової функції й завершується обчислювальний процес.

- існує принаймні один рядок з номером  $i = \overline{1, m}$ , для якого  $a_{ik}^q > 0$ . Відповідно до правила (3.11) визначається номер  $r$  стовпця, який буде виведеним з базису;

4. Перерахування елементів матриці  $\bar{A}$  й стовпця  $\bar{B}$  щодо нового базису відповідно до формул (3.12)-(3.14). Перехід до пункту 2 алгоритму.

**Таблична реалізація симплекс-методу.** З погляду забезпечення раціональності і наочності обчислювального процесу виконання алгоритму симплекс-методу зручно оформляти у вигляді послідовності таблиць. У різних джерелах наводяться різні модифікації симплекс-таблиць, що відрізняються одна від одної розташуванням окремих елементів. Однак всі вони базуються на тих самих принципах. Зупинимося на структурі таблиці, що наведена у таблиці 3.1:

Таблиця 3.1 – Структура симплекс-таблиці

Базис	$C_{\text{баз}}$	$C_j$	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$
		$B$	$A_1$	$A_2$	$A_j$	...	$A_n$
$A_j$							
$L_j$							
$\Delta_j$							

Стовпець «Базис» містить найменування базисних векторів (у тій послідовності, у якій вони входять до базису), стовпець  $C_{\text{баз}}$  – містить коефіцієнти цільової функції при базисних змінних, стовпець  $B$  - компоненти вектора обмежень щодо поточного базису,  $A_1$ - $A_n$  - компоненти матриці задачі щодо поточного базису. У рядку  $L_j$  записують індекси, визначені за формулою (3.9).

У рядку  $\Delta_j$  знаходяться поточні оцінки стовпців. Рядок  $C_j$  містить коефіцієнти при компонентах поточного плану в цільовій функції.

Слід зазначити, що таблична модифікація симплекс-методу має важливе практичне значення не стільки як зручна форма організації ручного розрахунку, скільки як основа для реалізації даного алгоритму в рамках програмного забезпечення ЕОМ.

Розглянемо приклад розв'язання ЗЛП симплекс-методом. Нехай дана канонічна задача ЛП:

$$\begin{aligned}
 L(x) &= 50x_1 - 10x_2 + 6x_3 + 40x_4 - 30x_5 \rightarrow \max \\
 x_1 + x_4 + 3x_5 &= 12, \\
 2x_1 + x_2 + 3x_4 &= 14, \\
 -2x_1 + 3x_3 - 4x_4 &= 17, \\
 x_1 > 0, x_2 > 0, x_3 > 0, x_4 > 0, x_5 > 0.
 \end{aligned}$$

Як видно, стовпці матриці з номерами  $\{5, 2, 3\}$  є лінійно незалежними. І можна одержати розкладання за цими стовпцями вектору обмежень із додатними коефіцієнтами. Останнє означає, що стовпці  $\{5, 2, 3\}$  утворюють припустимий базис, з якого можна почати розв'язання задачі. Початковий

опорний план має вигляд  $x^1 = \{0, 14, 17/3, 0, 4\}$ . Заповнюємо симплекс-таблицю, що відповідає першій ітерації ( $q=1$ ).

Таблиця 3.2 – Перша ітерація

Базис	$C_{баз}$	$C_j$	50	-10	6	40	-30
		$B$	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
$A_2$	-10	14	2	1	0	3	0
$A_3$	6	17/3	-2/3	0	1	-4/3	0
$A_5$	-30	4	1/3	0	0	1/3	1
$L_j$		-226	-34	-10	6	-48	-30
$\Delta_j$			-84	0	0	-88	0

Оскільки рядок оцінок у першому й четвертому стовпцях містить від'ємні елементи  $\Delta_1 = -84$ ,  $\Delta_4 = -88$ , план  $x^1 = \{0, 14, 17/3, 0, 4\}$  не є оптимальним, і значення цільової функції  $L(x^1) = -226$  можна покращити. Перейдемо до нового опорного плану.

Визначимо вектор, що будемо вводити до базису ( $A_1$  або  $A_4$ ).

Відповідно до алгоритму симплекс-методу визначимо відношення  $\Theta$ . Для вектора  $A_1$ :  $\Theta_r = \min_i \left\{ \frac{14}{2}; \frac{4}{1/3} \right\} = \min\{7; 12\} = 7$ ,  $r=2$ ; для вектора  $A_4$ :

$\Theta_r = \min_i \left\{ \frac{14}{3}; \frac{4}{1/3} \right\} = \min\{4,66; 12\} = 4,66$ ,  $r=2$ . Добутки для вектора  $A_1$ :  $-34 \cdot 7 = -238$ ;

для вектора  $A_4$ :  $-48 \cdot 4,66 = -223,7$ . Вважаємо номер стовпця, що вводиться до чергового базису,  $k = 1$  (тому що  $|-238| > |-223,7|$ ). З базису повинен бути виведений стовець із номером 2. Отримуємо черговий припустимий базис задачі  $\{1, 3, 5\}$ . Елемент, що перебуває на перетинанні виділених стовпця й рядка таблиці є ведучим, він дорівнює 2. Застосувавши формули (3.12)-(3.14), переходимо до симплекс-таблиці, що відповідає другій ітерації, і вважаємо індекс поточної ітерації  $q = 2$ .

Таблиця 3.3 – Друга ітерація

Базис	$C_{баз}$	$C_j$	50	-10	6	40	-30
		$B$	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
$A_1$	50	7	1	1/2	0	3/2	0
$A_3$	6	10,3	0	0,2	1	-0,3	0
$A_5$	-30	1,7	0	-0,2	0	-0,2	1
$L_j$		360,8	50	32,2	6	79,2	-30
$\Delta_j$			0	42,2	0	39,2	0

Одержуємо новий план  $x^2 = \{7; 0; 10,3; 0; 1,7\}$ . Як видно з таблиці, рядок оцінок містить тільки невід'ємні значення, тому досягнутий базис  $\{1, 3, 5\}$  є оптимальним. Таким чином, вектор  $x^* = \{7; 0; 10,3; 0; 1,7\}$  є оптимальним планом (точкою максимуму) задачі, максимальне значення цільової функції дорівнює  $L^* = L(x^*) = 360,8$ .

**Збіжність симплекс-методу. Виродженість у задачах ЛП.** Найважливішою властивістю будь-якого обчислювального алгоритму є



збіжність, тобто можливість одержання в ході його застосування шуканих результатів (із заданою точністю) за кінцеве число кроків (ітерацій).

Легко помітити, що проблеми зі збіжністю симплекс-методу потенційно можуть виникнути на етапі вибору значення  $r$  у випадку, коли однакові мінімальні значення відношення  $\Theta_r = \min_i \left\{ \frac{b_i}{a_{ij}} \right\}$  будуть досягнуті для кількох рядків таблиці одночасно. Тоді на наступній ітерації стовпець  $b^{q+1}$ -й буде містити нульові елементи. Нагадаємо, що припустимий базисний план канонічної задачі ЛП, що відповідає поточному базису, називається виродженим, якщо деякі його базисні компоненти дорівнюють нулю, тобто вектор  $b^q$  містить нульові елементи.

Задача ЛП, що має вироджені плани, називається виродженою. При виході на вироджений план ми фактично одержуємо розкладання вектора  $\bar{B}$  на меншу, ніж  $m$ , кількість векторів  $\bar{A}_j$  і, отже, втрачаємо можливість коректно визначити, введення якого стовпця до базису приведе до зростання значення цільової функції. Подібні ситуації, в остаточному підсумку, можуть привести до зациклення обчислювального процесу, тобто нескінченному перебору тих самих базисів.

З погляду першої геометричної інтерпретації ЗЛП ситуація виродженості означає, що через певну кутову точку багатогранної множини припустимих планів задачі, що відповідає поточному базисному плану, проходить більше ніж  $m$  гіперплощин обмежень задачі. Іншими словами, одне або кілька обмежень у даній точці є надлишковими. Останнє дає повід для міркувань про економічний зміст проблеми виродженості як проблеми наявності надлишкових обмежень і в деяких випадках дозволяє визначити шляхи її розв'язання.

Ідея методу розв'язання вироджених задач ЛП, що дістала назви **методу збурень**, полягає в тому, що при виході на вироджений план здійснюють незначний зсув вектора  $\bar{B}$ , і виродженість усувається.

Необхідно сказати, що розглянута проблема зациклення для більшості практично значущих задач є досить рідкою й малоймовірною. Більше того, вона дуже часто вирішується за рахунок помилок округлень при виконанні розрахунків на ЕОМ.

**Знаходження припустимого базисного плану.** У розглянутому вище прикладі вихідний базисний план, необхідний для початку обчислень за симплекс-методом, був підібраний за рахунок особливостей матриці умов. Дійсно, дана матриця вже містила необхідну кількість «майже базисних» стовпців. Очевидно, що для переважної більшості задач ЛП неможливо подібним чином відразу й у явному вигляді вказати вихідний припустимий базисний план. Існують різні прийоми розв'язання даної задачі. Ми зупинимось на одному з них, що отримав назву **методу мінімізації нев'язань**. Його сильною стороною є універсальність, хоча іноді він може опинитися занадто громіздким.

Ідея методу мінімізації нев'язань перебуває в побудові відповідної допоміжної задачі, для якої можна в явному вигляді вказати вихідний базисний план, і розв'язанні її за допомогою процедури симплекс-методу.

$$L = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \max \quad (3.15)$$

[illegible]

$$L = 0x_1 + 0x_2 + \dots + 0x_{n-1}x_{n+1} - \dots - 0 \cdot x_{n+m} \quad (3.17)$$

$$\begin{aligned} & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + 1^* x_{n+1} + \dots + 0^* x_{n+m} = b_1, \\ & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + 0^* x_{n+1} + 1^* x_{n+2} + \dots + 0^* x_{n+m} = b_2, \\ & \dots\dots\dots \\ & a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n + 0^* x_{n+1} + \overline{0^* x_{n+2}} + \dots + 1^* x_{n+m} = b_m, \end{aligned} \quad (3.18)$$

У випадку, коли оптимальний план допоміжної задачі всеж-таки містить не рівні нулю фіктивні компоненти, припустимі плани у вихідній задачі відсутні, тобто  $D = \emptyset$ .

### Контрольні запитання

1. Сформулюйте задачу лінійного програмування.
2. Дайте визначення для наступних понять: план, припустимий план, оптимальний план, розв'язок задачі.
3. Чим відрізняється загальна задача лінійного програмування від канонічної?
4. Чи завжди загальну задачу лінійного програмування можна привести до канонічного виду?
5. Яка точка опуклої множини називається кутовою?
6. У чому полягає перша геометрична інтерпретація задачі лінійного програмування?
7. Який план ЗЛП називається базисним?
8. Як зв'язані базисні плани й кутові точки області визначення задачі лінійного програмування?
9. Який план задачі лінійного програмування називається виродженим?
10. Сформулюйте критерій оптимальності припустимого базисного плану, застосовуваний у симплекс-методі.
11. Сформулюйте основні етапи стандартної ітерації симплекс-методу.
12. Для чого застосовується перетворення Жордана-Гауса?
13. Який елемент симплекс-таблиці називається ведучим?
14. При яких умовах робиться висновок про необмеженість цільової функції в розв'язуваній задачі?
15. Чи можна заздалегідь точно визначити кількість ітерацій, необхідних для розв'язання задачі симплекс-методом? Чи можна знайти верхню границю для даної величини?
16. Яка задача називається виродженою? За якими ознаками можна впізнати, що поточний план є виродженим?
17. У чому полягає основна ідея методу збурювань?
18. Для чого призначений метод мінімізації нев'язань?

## ТЕМА 4

# ТЕОРІЯ ДВОЇСТОСТІ І ДВОЇСТІ ОЦІНКИ В АНАЛІЗІ РОЗВ'ЯЗКІВ ЛІНІЙНИХ ОПТИМІЗАЦІЙНИХ МОДЕЛЕЙ

#### 4.1 Пряма і двоїста задачі як пара сполучених задач ЛП

Більшість задач лінійного програмування можна розглядати як економічні задачі про розподіл ресурсів  $b_1, b_2, \dots, b_m$  між різними споживачами, зокрема, між технологічними процесами  $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$ .

Будь-який припустимий розв'язок задачі лінійного програмування  $x_1, x_2, \dots, x_n$  дає конкретний розподіл ресурсів, що вказує частку кожного з них, яка буде використаною при реалізації певного технологічного процесу.

Загальна задача лінійного програмування формулюється в такий спосіб:  
знайти

$$L = c_1 x_1 + c_2 x_2 + \dots + c_n x_n \rightarrow \max, \quad (4.1)$$

де  $x$  задовольняє обмеженням

[illegible]

Тут  $a_{ij}$  – витрати  $i$ -го ресурсу на  $j$ -й технологічний процес. Зміст обмежень – сумарні витрати  $i$ -го ресурсу повинні не перевищувати кількість цього ресурсу, наявного в нашому розпорядженні  $b_i$ .

Для будь-якого технологічного процесу актуальним є також питання про те, щоб витрати ресурсів були мінімальними. Позначимо  $u_i$  ціну одиниці  $i$ -го ресурсу. Тоді можна записати вимогу, щоб вартість ресурсів була мінімальною

$$L' = b_1 u_1 + b_2 u_2 + \dots + b_m u_m \rightarrow \min,$$

а величина витрат на кожний вид технологічного процесу не має бути меншою за його вартість

$$\begin{aligned} a_{11}u_1 + a_{12}u_2 + \dots + a_{1m}u_m &\geq c_1, \\ a_{12}u_1 + a_{22}u_2 + \dots + a_{m2}u_m &\geq c_2, \\ &\dots\dots\dots \\ a_{1n}u_1 + a_{2n}u_2 + \dots + a_{mn}u_m &\geq c_n, \\ u_i &\geq 0. \end{aligned}$$

Таким чином, одержали задачу:

знайти такий розв'язок  $u^*$ , що перетворює на мінімум цільову функцію

$$L' = \sum_{i=1}^m b_i u_i \rightarrow \min \quad (4.3)$$

і задовольняє обмеженням

Двоїста

Знайти таке  $\bar{U}^*$ , що

$$\overline{B}^T \overline{U}^* \rightarrow \min$$

при обмеженнях

$$\overline{A}^T \overline{U} \geq \overline{C}, \quad \overline{U} \geq 0.$$

## 4.2 Основні теореми двоїстості, їхній економічний зміст

Фундаментальні властивості, якими володіють двоїсті задачі лінійного програмування, можуть бути сформульовані у вигляді стверджень, що приводяться нижче. Їх звичайно називають **теоремами подвійності**.

**Теорема 4.1.** (перша теорема подвійності). Якщо  $\bar{X}_0$  і  $\bar{U}_0$  - припустимі плани для пари двоїстих задач, тобто якщо

$$\bar{A} \bar{X}_0 \leq \bar{B} \quad \text{і} \quad \bar{A}^T \bar{U}_0 \geq \bar{C},$$

то

$$\bar{C}^T \bar{X}_0 \leq \bar{B}^T \bar{U}_0,$$

тобто значення цільової функції прямої задачі ніколи не перевищують значень цільової функції двоїстої задачі.

Доказ.

Оскільки  $\bar{U}_0$  - припустимий план, то

$$\bar{A}^T \bar{U}_0 \geq \bar{C}; \quad (4.5)$$

оскільки  $\bar{X}_0$  - припустимий план, то

$$\bar{A} \bar{X}_0 \leq \bar{B}. \quad (4.6)$$

Помножимо (4.5) на  $\bar{X}_0^T$

$$\bar{X}_0^T \bar{A}^T \bar{U}_0 \geq \bar{X}_0^T \bar{C}; \quad (4.7)$$

помножимо (4.6) на  $\bar{U}_0^T$

$$\bar{X}_0 \bar{A} \bar{U}_0^T \leq \bar{U}_0^T \bar{B} \quad (4.8)$$

і порівняємо (4.7) і (4.8). Оскільки

$$\bar{X}_0 \bar{A} \bar{U}_0^T = \left( \bar{X}_0^T \bar{A}^T \bar{U}_0 \right)^T, \quad \text{то} \quad \bar{U}_0^T \bar{B} \geq \bar{X}_0^T \bar{C},$$

або

$$\bar{U}_0^T \bar{B} \geq \bar{X}_0^T \bar{C}.$$

Зауваження. Теорема 4.1, зрозуміло, вірна й для оптимальних планів взаємно двоїстих задач:  $\bar{U}^* \bar{B}^T \geq \bar{X}^* \bar{C}^T$ , де  $\bar{X}^*$  і  $\bar{U}^*$  - будь-які оптимальні плани задач. Насправді, як буде видно з подальшого, справедлива рівність  $\bar{U}^* \bar{B}^T = \bar{X}^* \bar{C}^T$ .

**Теорема 4.2.** (друга теорема подвійності). Якщо для деяких припустимих планів  $\bar{X}_0$  і  $\bar{U}_0$  взаємно двоїстих задач виконується рівність

$$\bar{C}^T \bar{X}_0 = \bar{B}^T \bar{U}_0,$$

то  $\bar{X}_0$  і  $\bar{U}_0$  є оптимальними планами цих задач.

Доказ.

Відповідно до теореми 4 для всіх припустимих розв'язків  $\bar{X}$  і  $\bar{U}$  справедлива нерівність

$$\bar{C}^T \bar{X} \leq \bar{B}^T \bar{U},$$

але оскільки з умови теореми

$$\bar{C}^T \bar{X}_0 = \bar{B}^T \bar{U}_0,$$

то  $\bar{C}^T \bar{X}_0$  – найбільше з можливих значень цільової функції, тобто

$$\bar{C}^T \bar{X} \leq \bar{C}^T \bar{X}_0$$

цільова функція від  $\bar{X}_0$  – найбільша, таким чином,  $\bar{X}_0$  є оптимальним значенням.

Аналогічно для цільової функції двоїстої задачі

$$\bar{B}^T \bar{U}_0 \leq \bar{B}^T \bar{U},$$

тобто  $\bar{B}^T \bar{U}_0$  – найменше з можливих значень цільової функції двоїстої задачі, отже  $\bar{U}_0$  – оптимальне значення.

Таким чином, якщо для деяких припустимих планів  $\bar{X}_0$  і  $\bar{U}_0$  прямої і двоїстої задач їхні цільові функції рівні, то  $\bar{X}_0$  й  $\bar{U}_0$  – оптимальні плани пари сполучених задач.

**Теорема 4.3.** Якщо цільова функція в прямій задачі не обмежена зверху, то двоїста до неї задача не має припустимих планів.

Доказ.

Якщо припустити, що у двоїстій задачі існує хоча б один припустимий план  $\bar{U}_0$ , то відповідно до теореми 4.2, для будь-якого припустимого плану  $\bar{X}_0$  прямої задачі справедлива нерівність  $\bar{C}^T \bar{X} \leq \bar{B}^T \bar{U} < +\infty$ . Останнє означає, що цільова функція прямої задачі обмежена зверху. Оскільки це суперечить умові теореми, припущення про існування припустимих планів двоїстої задачі є невірним.

Наступне ствердження, відоме як **теорема рівноваги**, використовується при перевірці оптимальності планів ЗЛП.

**Теорема 4.4.** Нехай  $\bar{X}^*$  і  $\bar{U}^*$  – оптимальні плани канонічної та двоїстої стосовно неї задач. Якщо  $j$ -та компонента плану  $\bar{X}^*$  строго додатна ( $x_i^* > 0$ ), то відповідне  $j$ -е обмеження двоїстої задачі виконується як рівність:  $a_{1j}u_1 + \dots + a_{mj}u_m = c_j$ ; якщо ж  $j$ -та компонента плану  $\bar{X}^*$  має нульове значення ( $x_j^* = 0$ ), то  $j$ -е обмеження двоїстої задачі виконується як нерівність:  $a_{1j}u_1 + \dots + a_{mj}u_m > c_j$ .

Доказ.

Вектори  $\bar{X}^*$  і  $\bar{U}^*$  як припустимі плани, задовольняють обмеженням

$$\bar{A} \bar{X}^* = \bar{B}; \quad \bar{X}^* > 0 \text{ - прямої задачі}$$

$$\bar{A}^T \bar{U}^* - \bar{C}^T \geq 0; \quad \bar{U}^* > 0 \text{ - двоїстої задачі.}$$

Запишемо скалярний добуток

$$(\bar{A}^T \bar{U}^* - \bar{C}^T) \bar{X}^* = \bar{A}^T \bar{X}^* \bar{U}^* - \bar{C}^T \bar{X}^* = \bar{B}^T \bar{U}^* - \bar{C}^T \bar{X}^*.$$

Одержали різницю цільових функцій прямої і двоїстої задач. На підставі другої теореми подвійності оптимальні значення цільових функцій взаємно двоїстих задач збігаються. Отже скалярний добуток

$$(\bar{A}^T \bar{U}^* - \bar{C}^T) \bar{X}^* = 0.$$

Але скалярний добуток двох невід'ємних векторів може дорівнювати нулю тільки в тому випадку, якщо всі попарні добутки їхніх відповідних координат дорівнюють нулю. Тоді, якщо  $x_j > 0$ , то  $\sum_i a_{ij} u_i - c_j = 0$  або  $\sum_i a_{ij} u_i = c_j$ . А якщо  $x_j = 0$ , то можливо, що  $\sum_i a_{ij} u_i - c_j > 0$  або  $\sum_i a_{ij} u_i > c_j$ .

Практичне значення теорем подвійності полягає в тому, що вони дозволяють замінити процес розв'язання прямої задачі на розв'язання двоїстої, яке в певних випадках може виявитися більш простим. Наприклад, задача, область припустимих значень якої описується двома рівняннями, що зв'язують шість змінних ( $m = 2$ ,  $n = 6$ ), не може бути вирішеною графічним методом. Однак даний метод може бути застосований для розв'язання двоїстої до неї задачі, що має тільки дві змінні.

Оптимальний розв'язок двоїстої задачі можна одержати з таблиці, отриманої на фінальній ітерації процедури симплекс-методу. Елементи індексного рядка цієї таблиці  $L_j$  обчислюють відповідно до виразу (3.9)

$$L_j = \sum_{i=1}^m c_j a_{ij}, \text{ де } c_j - \text{елементи вектора-рядка, що містить коефіцієнти цільової}$$

функції прямої задачі при змінних, які є базисними в оптимальному плані;  $a_{ij}$  - елементи матриці  $\bar{D}^{-1}$ , що зворотна до матриці  $\bar{D}$ . Матриця  $\bar{D}$  складається з базисних векторів оптимального плану, компоненти яких узяті з початкового опорного плану задачі. Оптимальний план двоїстої задачі визначається співвідношенням

$$\bar{U}^* = \bar{c}_{j\text{баз}} \bar{D}^{-1}. \quad (4.9)$$

Зворотна матриця  $\bar{D}^{-1}$  завжди міститься в останній симплекс-таблиці у тих стовпчиках, де в першій таблиці задачі перебувала одинична матриця. (Нагадаємо, що добуток матриці на її зворотну дає одиничну матрицю, в якій діагональні елементи дорівнюють 1, а всі інші дорівнюють 0).

Таким чином, зв'язок між оптимальними розв'язками прямої й двоїстої задач і елементами індексних рядків  $L_j$  симплекс-таблиць, які відповідають цим розв'язкам, виражається наступними співвідношеннями:

$$\begin{aligned} a_{0,n+i}^{PP} &= u_i^*, \quad i = \overline{1, m}, \\ -a_{0,m+j}^{DB} &= x_j^*, \quad j = \overline{1, n}, \end{aligned}$$

де  $n$  – кількість змінних прямої задачі;

$m$  – кількість обмежень прямої задачі;



$a_{0,n+i}^{PP}$  -  $(n+i)$ -й елемент індексного рядка симплекс-таблиці прямої задачі, що містить оптимальний план;

$a_{0,m+j}^{DB}$  -  $(m+j)$ -й елемент індексного рядка симплекс-таблиці двоїстої задачі, що містить оптимальний план.

### 4.3 Двоїсті оцінки і дефіцитність ресурсів

У різних джерелах компоненти оптимального плану двоїстої задачі називають **двоїстими оцінками** або **тіньовими цінами**.

На основі теорем подвійності для пари задач ЛП у загальній формі можуть бути сформульовані деякі важливі з погляду економічної інтерпретації слідства. Зокрема, з теореми 4.4 випливає, що якщо при реалізації оптимального плану прямої задачі  $i$ -е обмеження виконується як строга нерівність, то оптимальне значення відповідної двоїстої змінної дорівнює нулю:

$$a_{i1}x_1^* + \dots + a_{in}x_n^* < b_i, \text{ то } u_i^* = 0.$$

Це означає, що якщо деякий ресурс  $b_i$  є в надлишковій кількості (не використовується повністю при реалізації оптимального плану), то  $i$ -е обмеження стає несуттєвим і тіньова оцінка такого ресурсу дорівнює нулю. Таким чином, тіньові оцінки характеризують **дефіцитність ресурсів**.

Якщо при реалізації оптимального плану двоїстої задачі  $j$ -е обмеження виконується як строга нерівність, то оптимальне значення відповідної змінної в оптимальному плані прямої задачі повинне дорівнювати нулю

$$a_{1j}u_1^* + \dots + a_{mj}u_m^* > c_j, \text{ то } x_j^* = 0.$$

Даний факт виражає **принцип рентабельності виробництва**. З огляду на економічний зміст двоїстих оцінок  $u_1^*, \dots, u_m^*$ , вираз  $a_{1j}u_1^* + \dots + a_{mj}u_m^*$  може бути інтерпретованим як питомі витрати на  $j$ -й технологічний процес. Отже, якщо ці витрати перевищують прибуток від реалізації одиниці  $j$ -го продукту, виробництво  $j$ -го продукту є нерентабельним і не повинне бути присутнім в оптимальному виробничому плані ( $x_j^* = 0$ ).

Незважаючи на можливі аналогії, які можуть виникнути у зв'язку з такими фундаментальними поняттями економічної теорії, як граничні витрати й граничний дохід, двоїсті оцінки не слід однозначно ототожнювати із цінами (хоча такі спроби іноді вживали на початковій стадії становлення дослідження операцій як науки).

### Контрольні запитання

1. Поясніть сутність подвійності в лінійному програмуванні.
2. Складіть просту економіко-математичну модель і запишіть до неї двоїсту. Дайте економічну інтерпретацію двоїстих оцінок.
3. Скільки змінних і обмежень має двоїста задача стосовно прямої задачі?
4. Поясніть економічний зміст першої теореми подвійності.
5. Поясніть економічний зміст другої теореми подвійності.
6. У чому полягає економічний зміст третьої теореми подвійності?

7. Сформулюйте правила побудови двоїстих задач.
8. Як, маючи оптимальний розв'язок прямої задачі, можна одержати оптимальний розв'язок двоїстої задачі?

## ТЕМА 5

### АНАЛІЗ ЛІНІЙНИХ МОДЕЛЕЙ ЕКОНОМІЧНИХ ЗАДАЧ

#### 5.1 Аналіз розв'язків лінійних економіко-математичних моделей

Традиційна економічна інтерпретація двоїстої задачі ЛП базується на моделі найпростішої задачі виробничого планування. У ній кожний  $j$ -й елемент вектора  $\bar{X}$  розглядається як план випуску продукції певного виду в натуральних одиницях,  $c_j$  - ціна одиниці продукції  $j$ -го виду,  $\bar{A}_j$  - вектор, що визначає технологію витрати наявних  $m$  ресурсів на виробництво одиниці продукції  $j$ -го виду,  $\bar{B}$  - вектор обмежень на обсяги цих ресурсів.

Припустимо, що для деяких значень  $\bar{A}_j$ ,  $\bar{B}$  і  $c_j$  знайдений оптимальний план  $x^*$  прямої задачі, що максимізує сумарний дохід  $\max_{x \in D} \{cx\} = cx^*$ , і визначені оптимальні оцінки сировини, тобто оптимальний план двоїстої задачі  $u^*$ . З виразу цільової функції двоїстої задачі  $L^* = b_1 u_1^* + b_2 u_2^* + \dots + b_m u_m^*$  видно, що величина двоїстої оцінки  $u_i^*$  показує, наскільки зросте максимальне значення цільової функції прямої задачі при збільшенні кількості сировини відповідного виду на одну одиницю. Таким чином, змінні двоїстої задачі  $u_1^*, \dots, u_m^*$  за своїм змістом є оцінками потенційної можливості одержання додаткового прибутку за рахунок збільшення обсягу відповідного ресурсу в умовах оптимального функціонування керованого економічного об'єкта.

Виникає питання про те, як буде змінюватися оптимальний план  $x^*$  при зміні компонентів вектора обмежень  $\bar{B}$  і, зокрема, при яких варіаціях  $\bar{B}$  оптимальний план  $x^*$  залишиться оптимальним. Дана задача одержала назву проблеми **стійкості оптимального плану**. Очевидно, що дослідження стійкості  $x^*$  має й безпосереднє практичне значення, тому що в реальному виробництві обсяги доступних ресурсів  $b_i$  можуть істотно коливатися після ухвалення планового розв'язку  $x^*$ .

Коли вектор обмежень  $\bar{B}$  отримує приріст  $\Delta b$ , виникають відповідні варіації для оптимального плану прямої задачі  $x^*(b+\Delta b)$  і значення цільової функції  $L[x^*(b+\Delta b)]$ . Допустимо, приріст  $\Delta b$  такий, що він не приводить до зміни оптимального базису задачі, тобто  $x^*(b+\Delta b) \geq 0$ . Уведемо функцію  $F(b)$ , що повертає оптимальне значення цільової функції задачі для різних значень вектора обмежень  $\bar{B}$

$$F(b) = \max_{x \in D(b)} \{cx\}. \quad (5.1)$$

Розглянемо відношення її приросту  $F(b+\Delta b)-F(b)$  до приросту аргументу  $\Delta b$ . Якщо для деякого  $i$  спрямувати  $\Delta b \rightarrow 0$ , одержимо

$$\lim_{\Delta b_i \rightarrow 0} \frac{F(b+\Delta b) - F(b)}{\Delta b_i} = \frac{\partial F(b)}{\partial b_i}. \quad (5.2)$$

З огляду на те, що відповідно до теореми 4.2 цільові функції пари сполучених задач при оптимальних планах дорівнюють одна одній, запишемо

$$F(b) = \sum_{j=1}^n c_j x_j^* = \sum_{i=1}^m b_i u_i^*. \quad (5.3)$$

Підставимо (5.3) до (5.2) і одержимо вираз

$$\frac{\partial F(b)}{\partial b_i} = \frac{\partial \left( \sum_{i=1}^m b_i u_i^* \right)}{\partial b_i} = u_i^*. \quad (5.4)$$

**Теорема 5.1.** В оптимальному плані двоїстої задачі значення змінної  $u_i^*$  чисельно дорівнює частинній похідній цільової функції  $L^*$  за відповідним аргументом  $b_i$ .

Звідси випливає економічна інтерпретація оптимальних змінних двоїстої задачі:

Кожний елемент  $u_1^*, \dots, u_m^*$  може розглядатися як гранична (миттєва) оцінка внеску  $i$ -го ресурсу в сумарний дохід  $L^*$  при оптимальному розв'язку  $x_1^*, \dots, x_n^*$ .

Інакше кажучи,  $u_i^*$  дорівнює приросту доходу, що виникає при збільшенні ресурсу  $i$  на одиницю за умови оптимального використання ресурсів.

Таким чином, якщо знайдено оптимальний план прямої задачі, можна провести аналіз стійкості двоїстих оцінок щодо змін компонентів вектора  $\bar{B}$ . Це дозволяє оцінити стійкість оптимального плану двоїстої задачі щодо зміни обмежень прямої задачі й ступінь впливу зміни  $\bar{B}$  на максимальне значення цільової функції, а також визначити найбільш доцільний варіант можливих змін  $\bar{B}$ .

План двоїстої задачі не змінюється для всіх значень  $b_i + \Delta b_i$ , при яких стовпець вектора  $\bar{B}^*$  останньої симплекс-таблиці не містить від'ємних чисел, тобто коли серед компонентів вектора

$$\bar{B}^* = \begin{bmatrix} b_1 + \Delta b_1 \\ b_2 + \Delta b_2 \\ \dots\dots\dots \\ b_m + \Delta b_m \end{bmatrix} \quad (5.5)$$

немає від'ємних.  $\bar{B}^*$  – матриця, складена з компонентів векторів базису, що визначає оптимальний план задачі, тому що базисні компоненти оптимального плану перебувають у стовпці вектора  $\bar{B}$  останньої симплекс-таблиці.

Елементи  $(n+i)$ -го стовпця  $a_{ij}$  останньої симплекс-таблиці, що містить оптимальну оцінку  $i$ -го ресурсу  $u_i^*$ , показують, на скільки одиниць зміняться

$$\overline{B}^* = \left| \begin{array}{l} x_1 + a_{1,n+i}\Delta b_i \geq 0 \\ x_2 + a_{2,n+i}\Delta b_i \geq 0 \\ ..... \\ x_n + a_{n,n+i}\Delta b_i \geq 0 \end{array} \right|. \quad (5.6)$$

Таблиця 5.1 – Симплекс-таблиця, що містить оптимальний план

Звідки оптимальний план прямої задачі  $x^*=(83; 36; 0; 0; 21)$ , оптимальний план двоїстої задачі  $u^*=(\frac{9}{7}; \frac{2}{7}; 0)$ . Більш дефіцитною є сировина  $S_1$ , тому що її тіньова оцінка вище й відповідно сильніше впливає на величину цільової функції.

якщо виробництво виробів  $A$  збільшити на  $\frac{4}{7}$  одиниці, а виробництво виробів  $B$  знизити на  $\frac{3}{7}$  одиниці, при цьому витрата сировини  $S_3$  зросте на  $\frac{11}{28}$  одиниці.

Визначимо інтервали стійкості двоїстих оцінок. Для ресурсу 1 відповідно до елементів стовпчика  $A_3$  маємо

Запишемо вектор  $\bar{B}^*$  з умовами його невід'ємності й визначимо межі припустимих значень  $\Delta b_1$

44

Оптимальний план двоїстої задачі залишиться незмінним, якщо  $\Delta b_1$  належить інтервалу  $-53,8 \leq \Delta b_1 \leq 83,7$ , а перший ресурс  $440 - 53,7 \leq b_1 \leq 440 + 83,7$  або  $386,3 \leq b_1 \leq 523,7$ .

Аналогічно для ресурсу 2 відповідно до елементів стовпчика  $A_4$  запишемо

$$x^* = (83 - 0,43\Delta b_2; 36 + 0,57\Delta b_2; 0; 0; 21 - 1,57\Delta b_2).$$

$$\bar{B}^* = \left| \begin{array}{l} 83 - 0,43\Delta b_2 \geq 0 \\ 36 + 0,57\Delta b_2 \geq 0 \\ 21 - 1,57\Delta b_2 \geq 0 \end{array} \right| = \left| \begin{array}{l} \Delta b_2 \leq 193 \\ \Delta b_2 \geq -63,2 \\ \Delta b_2 \leq 13,4 \end{array} \right|$$

Оптимальний план двоїстої задачі залишиться незмінним, якщо  $\Delta b_2$  належить інтервалу  $-63,2 \leq \Delta b_2 \leq 13,4$ , а другий ресурс  $393 - 63,2 \leq b_2 \leq 393 + 13,4$  або  $329,8 \leq b_2 \leq 406,4$ .

Таким чином, якщо збільшення кількості ресурсів  $S_1$  належить проміжку  $-53,8 < \Delta b_1 < 83,7$ , а кількість інших ресурсів незмінна, або збільшення кількості ресурсів  $S_2$  належить проміжку  $-63,2 < \Delta b_2 < 83,7$ , а кількість інших ресурсів незмінна, то двоїста задача має той самий оптимальний план  $u^* = (1,286; 0,286; 0)$ .

Стосовно прямої задачі, можна показати, що при зміні кількості першого ресурсу  $S_1$  у межах  $386,3 \leq b_1 \leq 523,7$  можливий дохід підприємства лежить у межах  $609,7 \leq L^* \leq 784,3$  а оптимальний план прямої задачі

$$(52,3; 59,1; 0; 0; 0,02) \leq x^* \leq (131; 0; 0; 0; 53,6).$$

При зміні кількості другого ресурсу  $S_2$  у межах  $329,8 \leq b_2 \leq 406,4$  можливий дохід підприємства лежить у межах  $661 \leq L^* \leq 681,6$  а оптимальний план прямої задачі  $(110; 0; 0; 0; 53,6) \leq x^* \leq (77; 43,6; 0; 0; 0)$ .

Розраховані інтервали стосуються випадків зміни тільки одного ресурсу. У випадку одночасної зміни обсягу всіх або кількох ресурсів для визначення нового оптимального плану використовують одне з основних співвідношень обчислювальної процедури симплекс-методу:

$$x^* = \bar{D}^{-1} * \bar{B}, \quad (5.7)$$

де  $\bar{D}$  - матриця, що складається з базисних векторів оптимального плану, компоненти яких узяті з початкового опорного плану;

$\bar{B}$  - вектор обмежень.

У розглянутому числовому прикладі матриці  $\bar{D}$  й  $\bar{D}^{-1}$  відповідно мають вигляд

$$\bar{D} = \left| \begin{array}{ccc} 4 & 3 & 0 \\ 3 & 4 & 0 \\ 3 & 5 & 1 \end{array} \right|; \quad \bar{D}^{-1} = \left| \begin{array}{ccc} 4/7 & -3/7 & 0 \\ -3/7 & 4/7 & 0 \\ 11/28 & -11/7 & 1 \end{array} \right|$$

Нехай новий вектор обмежень

$$\bar{B} = \left| \begin{array}{c} 440+84 \\ 393+13,4 \\ 450+0 \end{array} \right| = \left| \begin{array}{c} 524 \\ 406,4 \\ 450 \end{array} \right|$$

тоді новий оптимальний план визначиться в такий спосіб

$$x^* = \begin{vmatrix} 4/7 & -3/7 & 0 \\ -3/7 & 4/7 & 0 \\ 11/28 & -11/7 & 1 \end{vmatrix} * \begin{vmatrix} 524 \\ 406,4 \\ 450 \end{vmatrix} = \begin{vmatrix} 125,26 \\ 7,66 \\ 17,2 \end{vmatrix}$$

Тобто  $x^*=(125,26; 7,66; 0; 0; 17,2)$ , при якому прибуток дорівнюватиме 790 грош. од.

## 5.2 Аналіз параметричної стійкості розв'язків ЗЛП

З погляду економічної інтерпретації задачау дослідження параметричної стійкості можна розглядати як вивчення тих меж коливання цін на продукцію керованого підприємства (фірми), при яких прийнятий план випуску продукції продовжує залишатися оптимальним. Таким чином, питання стійкості оптимального плану ЗЛП може бути поставленим для випадку варіації коефіцієнтів цільової функції

$$c_j, \quad j = \overline{1, n}.$$

Зміст проблеми стійкості оптимального плану ЗЛП стосовно варіацій цільової функції можна проілюструвати за допомогою першої геометричної інтерпретації. На рисунку 5.1 зображено множину припустимих планів  $D$  певної задачі ЛП. Як видно з рисунку, цільова функція  $L$  досягає екстремального значення в точці  $x^*$ , а зміні її коефіцієнтів від  $c$  до  $c'$  або  $c''$  на

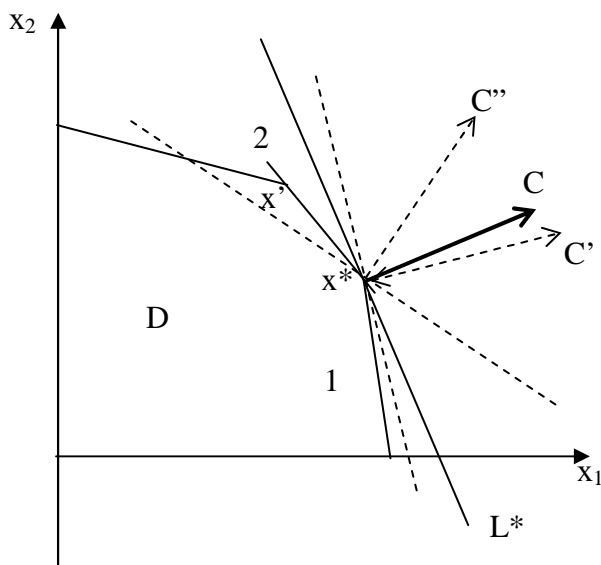


Рисунок 5.1 - Графічна інтерпретація параметричної стійкості оптимального плану

умови оптимальності плану ЗЛП

рисунку відповідає поворот лінії рівня відносно  $x^*$ . Активним обмеженням (тобто таким, що звертаються на рівність) у точці  $x^*$  відповідають лінії 1 і 2. Доти, поки при повороті, викликаному зміною вектора  $c$ , лінія рівня цільової функції не виходить за межі утвореної лініями обмежень множини,  $x^*$  залишається оптимальним планом. Як показано на рисунку 5.1, цей план не змінюється при переході від  $c$  до  $c'$ , і, навпаки, при переході від  $c$  до  $c''$  лінія рівня цільової функції  $L(x)=c''x$  перетинає лінію 2, що викличе зміну оптимального базисного плану, яким тепер стане точка  $x'$ . Використовуючи

$$\Delta_i = L_i - c_i \geq 0, \quad (5.8)$$

можна одержати кількісні оцінки для меж коливань коефіцієнтів цільової функції, при яких не відбувається зміна оптимального плану. Припустимо, що варіації піддався деякий елемент  $c_r' = c_r + \Delta c_r$ .

Можливі два випадки:

1. Стовець  $r$  не входить до оптимального базису. Тоді для незмінності оптимального плану необхідно й достатньо виконання умови

$$\Delta_r' = L_r - c_r' \geq 0.$$

Звідси можна одержати значення для припустимої варіації

$$\Delta c_r \leq L_r - c_r. \quad (5.9)$$

2. Стовець  $r$  входить до оптимального базису. У цьому випадку для збереження оптимальності поточного плану буде потрібне виконання для всіх небазисних стовпців умов (5.8) або

$$\Delta_j = \sum_{i=1}^m c_{j_{\text{баз}}} a_{ij} - c_j \geq 0, \quad (5.10)$$

оскільки у цьому випадку зміни відбуваються також у стовпчику  $C_{\text{баз}}$  симплекс-таблиці, а це, у свою чергу, стосується всіх ненульових оцінок  $\Delta_j$ .

Отже, у цьому випадку припустима варіація повинна задовольняти умовам

$$\Delta c_r \leq \sum_{i=1}^m c_{j_{\text{баз}}} a_{ij} - c_j. \quad (5.11)$$

Повернемося до числового прикладу й визначимо межі зміни параметрів цільової функції, при яких знайдений план  $x^*=(83; 36; 0; 0; 21)$  залишається оптимальним. У даній задачі інтерес представляють варіації коефіцієнтів  $c_1$  і  $c_2$ , які стоять при базисних змінних в оптимальному плані.

Запишемо умови (5.11) для коефіцієнта  $c_1$

$$\begin{cases} \Delta_3 = L_3 - c_3 = (6 + \Delta c_1) * 4/7 + 5 * (-3/7) + 0 * 11/28 - 0 = \frac{24}{7} + \frac{4}{7} \Delta c_1 - \frac{15}{7} \\ \Delta_4 = L_4 - c_4 = (6 + \Delta c_1) * (-3/7) + 5 * 4/7 + 0 * 11/7 - 0 = -\frac{18}{7} - \frac{3}{7} \Delta c_1 + \frac{20}{7} \end{cases} \Rightarrow$$

$$\begin{cases} \Delta_3 = \frac{24}{7} + \frac{4}{7} \Delta c_1 - \frac{15}{7} \geq 0 \\ \Delta_4 = -\frac{18}{7} - \frac{3}{7} \Delta c_1 + \frac{20}{7} \geq 0 \end{cases} \Rightarrow \begin{cases} \Delta c_1 \geq -\frac{9}{4} \\ \Delta c_1 \leq \frac{2}{3} \end{cases},$$

$$-\frac{9}{4} \leq \Delta c_1 \leq \frac{2}{3},$$

$$3,75 \leq c_1 \leq 6,67.$$

Аналогічно визначимо варіацію коефіцієнта  $c_2$ .

$$\begin{cases} \Delta_3 = L_3 - c_3 = 6 * 4/7 + (5 + \Delta c_2) * (-3/7) + 0 * 11/28 - 0 = \frac{9}{7} - \frac{3}{7} \Delta c_2 \\ \Delta_4 = L_4 - c_4 = 6 * (-3/7) + (5 + \Delta c_2) * 4/7 + 0 * (-11/7) - 0 = \frac{2}{7} + \frac{4}{7} \Delta c_2 \end{cases} \Rightarrow$$

$$\begin{cases} \Delta_3 = \frac{9}{7} - \frac{3}{7}\Delta c_2 \geq 0 \\ \Delta_4 = -\frac{2}{7} + \frac{4}{7}\Delta c_2 \geq 0 \end{cases} \Rightarrow \begin{cases} \Delta c_2 \leq 3 \\ \Delta c_2 \geq \frac{1}{2} \end{cases},$$

$$\frac{1}{2} \leq \Delta c_2 \leq 3,$$

$$5,5 \leq c_2 \leq 8.$$

Наведений приклад дослідження чутливості оптимального плану стосовно зміни параметрів задачі є простим. Існують і більш складні задачі, у яких, наприклад, досліджуються спільні варіації параметрів різних типів. Вони складають предмет спеціального розділу дослідження операцій, що одержав назву параметричного програмування.

### 5.3 Оцінка рентабельності виробленої продукції

Оцінка рентабельності продукції, що випускається підприємством, ґрунтується на теоремі рівноваги, (теоремі 4.4) і виконується за допомогою двоїстих оцінок і обмежень двоїстої задачі, що характеризують кожний вид продукції. Із цієї теореми випливає, що кількість продукції, витрати на виробництво якої перевищують дохід, в оптимальному плані дорівнює нулю.

Якщо при підстановці  $u^*$  до системи обмежень двоїстої задачі вартість ресурсів, витрачених на одиницю продукції (ліва частина), перевищує ціну цієї продукції (права частина), то виробництво такої продукції для підприємства недоцільно. Тобто у цьому випадку даний вид продукції є нерентабельним. Якщо ж співвідношення виконується як рівність, продукція рентабельна.

Аналогічні результати можна одержати, проаналізувавши симплекс-різниці  $\Delta_j$  у стовпчиках, що відповідають досліджуваним видам продукції. Їхні значення показують, наскільки вартість ресурсів перевищує ціну одиниці відповідної продукції. Так, якщо симплекс-різниця дорівнює нулю  $\Delta_j = 0$ , то продукція рентабельна. І, навпаки, якщо  $\Delta_j \neq 0$ , то відповідна продукція нерентабельна. Помітимо також, що індексний рядок останньої симплекс-таблиці містить значення додаткових змінних двоїстої задачі  $u^*$ . Таким чином, якщо вони перевищують ціну продукції відповідного виду, то цей вид продукції є рентабельним.

Звернемося до числового приклада. Підставимо отримані тіньові оцінки в обмеження двоїстої задачі:

$$\begin{cases} 4 * \frac{9}{7} + 3 * \frac{2}{7} + 3 * 0 = 6 & \text{продукція A рентабельна,} \\ 3 * \frac{9}{7} + 4 * \frac{2}{7} + 5 * 0 = 5 & \text{продукція B рентабельна.} \end{cases}$$

Оскільки обоє обмеження виконуються як строгі рівності, обидва види продукції A і B є рентабельними. Це підтверджується і тим, що в оптимальному



плані  $x^*=(83; 36; 0; 0; 21)$  обидва елементи  $x_1$  і  $x_2$ , що відповідають обсягам виробництва, не нульові.

Проаналізуємо додаткові змінні двоїстої задачі, які розміщуються в індексному рядку симплекс-таблиці в стовпчиках  $A_1$  і  $A_2$ . Їхні оптимальні значення  $u_4=6$ ;  $u_5 = 5$ . Відповідно симплекс-різниці  $\Delta_1 = \Delta_2 = 0$ , що також свідчить про рентабельність продукції А і В.

#### 5.4 Аналіз обмежень дефіцитних і недефіцитних ресурсів

Статус ресурсів прямої задачі можна визначити за трьома способами. Перший - підстановкою  $x^*$  до системи обмежень прямої задачі. Якщо обмеження виконується як рівність, то відповідний ресурс є дефіцитним, у протилежному випадку - недефіцитним.

Другий спосіб - за допомогою додаткових змінних прямої задачі. Якщо додаткова змінна в оптимальному плані дорівнює нулю, то відповідний ресурс дефіцитний, а якщо відмінна від нуля - ресурс недефіцитний.

Третій спосіб - за допомогою двоїстих оцінок. Якщо  $u_i \neq 0$ , то зміна (збільшення або зменшення) обсягів  $i$ -го ресурсу приводить до відповідної зміни доходу підприємства, і тому такий ресурс є дефіцитним. Якщо  $u_i=0$ , то  $i$ -й ресурс недефіцитний.

Звернемося до числового прикладу. Підставимо компоненти оптимального плану  $x^*=(83; 36; 0; 0; 21)$  до системи обмежень прямої задачі:

$$\begin{cases} 4 \cdot 83 + 3 \cdot 36 = 440 & \text{ресурс витрачено повністю,} \\ 3 \cdot 83 + 4 \cdot 36 = 393 & \text{ресурс витрачено повністю,} \\ 3 \cdot 83 + 5 \cdot 36 = 429 & \text{ресурс витрачено не повністю.} \end{cases}$$

Змінні  $u_1 = 1,286$  і  $u_2 = 0,286$  є умовними двоїстими оцінками одиниці сировини  $S_1$  і  $S_2$  відповідно. Ці оцінки відмінні від нуля, а сировина  $S_1$  і  $S_2$  повністю використовується при оптимальному плані виробництва виробів А і В. Двоїста оцінка одиниці сировини  $S_3$   $u_3=0$ . Цей вид сировини не використовується повністю при оптимальному плані виробництва.

Таким чином, додатну двоїсту оцінку мають лише ті види сировини, які повністю використовуються при оптимальному плані виробництва. Тому двоїсті оцінки визначають **дефіцитність** використовуваної підприємством сировини.

$$\begin{cases} u_1 = 1,286 & \text{ресурс 1 дефіцитний,} \\ u_2 = 0,286 & \text{ресурс 2 дефіцитний,} \\ u_3 = 0 & \text{ресурс 3 недефіцитний.} \end{cases}$$

#### Контрольні запитання

1. У чому полягає економічна інтерпретація прямої та двоїстої задач лінійного програмування?
2. Як визначити, є ресурс дефіцитним чи ні?
3. Як визначити, що продукція є рентабельною або нерентабельною?

4. У чому полягає економічний зміст змінних двоїстої задачі?
5. Який зміст вкладають у поняття «параметрична стійкість»?
6. Сформулюйте умови для припустимих змін цільової функції задачі, при яких її оптимальний план залишається незмінним.
7. Як визначити статус ресурсів прямої задачі?
8. Як визначити інтервали стійкості двоїстих оцінок щодо зміни запасів дефіцитних ресурсів?
9. Як визначити оптимальний план виробництва продукції й зміну доходу підприємства при збільшенні або зменшенні обсягу ресурсів?
10. Як розрахувати інтервали можливої зміни ціни одиниці кожного виду продукції?

## ТЕМА 6 ТРАНСПОРТНА ЗАДАЧА

### 6.1 Транспортна задача в матричній постановці і її властивості

**Економіко-математична модель задачі.** Нехай є  $m$  пунктів відправлення (постачальників)  $A_1, A_2, \dots, A_m$ , у яких знаходиться однорідна продукція в кількостях  $a_1, a_2, \dots, a_m$  відповідно. Нехай є  $n$  пунктів призначення (споживачів)  $B_1, B_2, \dots, B_n$ , що подали заявки відповідно на  $b_1, b_2, \dots, b_n$  одиниць вантажу. Сума всіх заявок дорівнює сумі всіх запасів.

Відомі вартості  $c_{ij}$  перевезення одиниці продукції з  $i$ -го пункту відправлення в  $j$ -й пункт призначення. Вважається, що вартість перевезення кількох одиниць вантажу пропорційна їхньому числу.

Потрібно скласти такий план перевезень (звідки, куди й скільки одиниць везти), щоб всі заявки були виконані, а загальна вартість всіх перевезень була мінімальною.

Поставимо цю задачу як задачу лінійного програмування. Позначимо  $x_{ij}$  кількість одиниць вантажу, що відправляється з  $i$ -го ПВ  $A$  до  $j$ -го ПП  $B$ .

Сукупність чисел  $x_{ij}$  будемо називати «планом перевезень», а самі величини  $x_{ij}$  – «перевезеннями». Ці невід’ємні змінні повинні задовольняти наступним умовам:

1. Сумарна кількість вантажу, що направляється з кожного ПВ до усіх ПП, повинна дорівнювати запасу вантажу в даному пункті.
2. Сумарна кількість вантажу, що доставляється до кожного ПП із усіх ПВ, повинна дорівнювати заявці, поданої даним пунктом.
3. Сумарна вартість всіх перевезень, тобто сума величин  $x_{ij}$ , помножених на відповідні вартості  $c_{ij}$ , повинна бути мінімальною.

Таким чином, задача зводиться до визначення такого плану перевезень деякого продукту з пунктів його виробництва до пунктів споживання  $x = (x_{11}, \dots, x_{1n}, x_{21}, \dots, x_{2n}, \dots, x_{i1}, \dots, x_{in}, \dots, x_{m1}, \dots, x_{mn})$ , який мінімізує цільову функцію

$$f(x) = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \quad (6.1)$$

на множині припустимих планів

$$\sum_{j=1}^n x_{ij} = a_i; \sum_{i=1}^m x_{ij} = b_j; x_{ij} \geq 0; i = \overline{1, m}; j = \overline{1, n} \quad (6.2)$$

при дотриманні умови балансу

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j. \quad (6.3)$$

Якщо умова (6.3) виконується, то задача називається **збалансованою** або **закритою**, а інакше задача називається **незбалансованою** або **відкритою**.

Транспортна задача є представником класу задач лінійного програмування й тому має всі якості лінійних оптимізаційних задач, але одночасно вона має й ряд додаткових корисних властивостей, які дозволили розробити спеціальні методи її розв'язання.

Число лінійно незалежних серед рівнянь в умовах-обмеженнях ТЗ дорівнює

$$m + n - 1,$$

отже число базисних змінних так само дорівнює

$$m + n - 1.$$

Загальне число змінних  $x_{ij}$  у транспортній задачі дорівнює  $m * n$ , а число вільних змінних

$$k = mn - (m + n - 1) = (m - 1)(n - 1).$$

Відомо, що в задачі лінійного програмування оптимальний розв'язок досягається в одній з вершин ОДР, в опорній точці, де принаймні  $k$  змінних дорівнюють нулю. Виходить, у нашому випадку для оптимального плану принаймні  $(m-1)(n-1)$  перевезень повинні дорівнювати нулю (з відповідних ПВ у відповідні ПП нічого не перевозиться).

Будемо називати будь-який план перевезень **припустимим**, якщо він задовольняє умовам-обмеженням (всі заявки задоволені, всі запаси вичерпані).

Припустимий план будемо називати **опорним**, якщо в ньому відмінні від нуля не більше  $m + n - 1$  базисних перевезень, а інші  $(m-1)(n-1)$  перевезень дорівнюють нулю.

План будемо називати **оптимальним**, якщо він, серед всіх припустимих планів, приводить до мінімальної сумарної вартості перевезень ( $L = \min$ ).

У силу особливої структури ТЗ при її розв'язанні не приходится довго розв'язувати систему рівнянь. Всі операції із знаходження оптимального плану зводяться до маніпуляцій безпосередньо з таблицею, де в певному порядку записані умови транспортної задачі: перелік ПВ й ПП, заявки й запаси, а також вартості перевезень  $c_{ij}$ . У міру заповнення цієї таблиці в її клітках проставляють самі перевезення  $x_{ij}$ . Транспортна таблиця складається з  $m$  рядків і  $n$  стовпців. Рядки транспортної таблиці відповідають пунктам виробництва (в останній клітці кожного рядка зазначений обсяг запасу продукту  $a_i$ ), а стовпці - пунктам споживання (остання клітка кожного стовпця містить значення заявки  $b_j$ ). У правому верхньому куті кожної клітки ставлять вартість  $c_{ij}$  перевезення одиниці продукту з  $A_i$  до  $B_j$ , а центр клітки залишають вільним, щоб поміщати у нього саме перевезення  $x_{ij}$ . Клітки, які містять нульові перевезення ( $x_{ij}=0$ ), називають вільними, а ненульові - зайнятими ( $x_{ij}>0$ ).

## 6.2 Методи побудови опорного плану

За аналогією з іншими задачами лінійного програмування розв'язання транспортної задачі починається з побудови припустимого базисного плану. Найбільш простий спосіб його знаходження ґрунтується на так званому **методі північно-західного кута**. Суть методу полягає в послідовному розподілі всіх запасів, наявних у першому, другому і т.д. пунктах виробництва, до першого, другого і т.д. пунктів споживання. Кожний крок розподілу зводиться до спроби повного вичерпання запасів у черговому пункті виробництва або до спроби повного задоволення потреб у черговому пункті споживання. На кожному кроці  $q$  величини поточних нерозподілених запасів позначаються  $a_i^q$ , а поточних незадоволених потреб -  $b_j^q$ . Побудова припустимого початкового плану, відповідно до методу північно-західного кута, починається з лівого верхнього кута транспортної таблиці. Для чергової клітки, розташованої в рядку  $i$  і стовпці  $j$ , розглядаються значення нерозподіленого запасу в  $i$ -ому пункті виробництва й незадоволеної заявки в  $j$ -ому пункті призначення, з них вибирається мінімальне й призначається як обсяг перевезення між даними пунктами:  $x_{ij} = \min\{a_i^q, b_j^q\}$ . У результаті цього значення нерозподіленого запасу й незадоволеної потреби у відповідних пунктах зменшуються:

$$a_i^{(q+1)} = a_i^q - x_{ij}; \quad b_j^{(q+1)} = b_j^q - x_{ij}.$$

Очевидно, що на кожному кроці виконується хоча б одна з рівностей:  $a_i^{(q+1)} = 0$  або  $b_j^{(q+1)} = 0$ . Якщо справедливо  $a_i^{(q+1)} = 0$ , то це означає, що весь запас  $i$ -го пункту виробництва вичерпаний і необхідно перейти до розподілу запасу в пункті виробництва  $i + 1$ , тобто переміститися до наступної клітки вниз по стовпцю. Якщо ж  $b_j^{(q+1)} = 0$ , то значить повністю задоволена заявка для  $j$ -го пункту, після чого виконується перехід на клітку, розташовану праворуч по рядку. Знову обрана клітка стає поточною, і для неї повторюються всі перелічені операції.

Ґрунтуючись на умові балансу запасів і заявок (6.3), неважко довести, що за кінцеве число кроків буде отриманий припустимий план. У силу тієї ж умови число кроків алгоритму не може бути більшим за  $m+n-1$ , тому завжди залишаться вільними (нульовими)  $mn-(m+n-1)$  кліток. Отже, отриманий план є базисним. Не виключено, що на деякому проміжному кроці поточний нерозподілений запас виявиться рівним поточній незадоволеній заявці ( $a_i^q = b_j^q$ ). У цьому випадку перехід до наступної клітки відбувається в діагональному напрямку (одночасно змінюються поточні пункти виробництва й призначення), а це означає «втрату» одного ненульового компонента в плані або, інакше кажучи, виродженість побудованого плану.

Розглянемо приклад. З 3-х пунктів виробництва необхідно вивезти однорідний продукт в 5 пунктів споживання. Транспортні витрати, обсяг виробництва й споживання наведені в таблиці 6.1.

Таблиця 6.1 - Вихідні дані

	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>B<sub>5</sub></b>	<b>Запаси</b>
<b>A<sub>1</sub></b>	7	5	2	8	7	125
<b>A<sub>2</sub></b>	8	9	4	6	9	60
<b>A<sub>3</sub></b>	5	1	9	2	3	115
<b>Заявки</b>	30	50	100	40	80	300

Зауважимо, що запаси дорівнюють заявкам. Таким чином, задача є збалансованою.

Визначимо кількість базисних змінних

$$m + n - 1 = 3 + 5 - 1 = 7.$$

Кількість вільних змінних

$$(n-1)*(m-1) = (5-1)*(3-1) = 8.$$

Іншими методами визначення початкового опорного плану є метод мінімальної вартості, метод подвійної переваги або метод апроксимації Фогеля.

У таблиці 6.2 показаний процес пошуку припустимого плану за **методом північно-західного кута**, включаючи послідовну зміну обсягу нерозподілених запасів і незадоволених потреб. Стрілки відображають траєкторію переходу по клітках транспортної таблиці, а цифри, що знаходяться за її межами, - поточні нерозподілені залишки після призначення обсягу для чергової клітки.

Таблиця 6.2 - Визначення опорного плану за методом північно-західного кута

	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>B<sub>5</sub></b>	<b>Запаси</b>			
<b>A<sub>1</sub></b>	30 ⇒	50 ⇒	45 ↓			125	95	45	0
<b>A<sub>2</sub></b>			55 ⇒	5 ↓		60		5	0
<b>A<sub>3</sub></b>				35 ⇒	80	115		80	0
<b>Заявки</b>	30	50	100	40	80	300			
	0	0	55						
			0	35					
				0	0				

Знайдений опорний план  $x = (30, 50, 45, 0, 0, \dots)$ . Значення цільової функції при цьому плані перевезень

$$L = 30*7+50*5+45*2+55*4+5*6+35*2+80*3 = 1110.$$

Особливістю припустимого плану, побудованого за методом північно-західного кута, є те, що цільова функція на ньому приймає значення, як правило, далеко від оптимального. Це відбувається тому, що при його побудові ніяк не враховуються значення  $c_{ij}$ . У зв'язку із цим на практиці для одержання вихідного плану використовується інший спосіб - метод мінімального елемента, у якому при розподілі обсягів перевезень у першу чергу займаються клітки з найменшими цінами.

**Метод мінімальної вартості** полягає в тому, що в таблиці вартостей вибирають найменшу, і в цій клітці записують найменше із чисел  $(a_i, b_j)$ , таблиця 6.3.

Таблиця 6.3 - Визначення опорного плану за методом мінімальної вартості

	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>B<sub>5</sub></b>	<b>Запаси</b>
<b>A<sub>1</sub></b>	7 <b>25</b>	5	2 <b>100</b>	8	7	125
<b>A<sub>2</sub></b>	8 <b>5</b>	9	4	6	9 <b>55</b>	60
<b>A<sub>3</sub></b>	5	1 <b>50</b>	9	2 <b>40</b>	3 <b>25</b>	115
<b>Потреби</b>	30	50	100	40	80	300

Значення цільової функції при цьому плані перевезень

$$L = 25 \cdot 7 + 100 \cdot 2 + 5 \cdot 8 + 55 \cdot 9 + 50 \cdot 1 + 40 \cdot 2 + 25 \cdot 3 = 1115.$$

**Метод подвійної переваги** показаний у таблиці 6.4. Він полягає в тому, що в кожному стовпці відзначають знаком **V** клітку з найменшою вартістю, потім те ж роблять у кожному рядку. У клітки з **W** заносять самі більші обсяги перевезень. Потім розподіляють перевезення по клітках, відзначеним **V**. У частині таблиці, що залишилася, перевезення розподіляють за методом найменшої вартості.

Таблиця 6.4. - Визначення опорного плану за методом подвійної переваги

	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>B<sub>5</sub></b>	<b>Запаси</b>
<b>A<sub>1</sub></b>	7 <b>25</b>	5	2 <b>100</b> <b>W</b>	8	7	125
<b>A<sub>2</sub></b>	8 <b>5</b>	9	4 <b>V</b>	6	9 <b>55</b>	60
<b>A<sub>3</sub></b>	5 <b>V</b>	1 <b>50</b> <b>W</b>	9	2 <b>40</b> <b>V</b>	3 <b>25</b> <b>V</b>	115
<b>Заявки</b>	30	50	100	40	80	300

Отриманий опорний план збігається із планом за методом мінімальної вартості,  $L = 1115$ .

Приймати в якості опорного треба той план, для якого транспортні витрати виявилися найменшими. Таким чином, у якості опорного треба прийняти план, отриманий за методом північно-західного кута.

Цей план є припустимим, тому що суми за рядками дорівнюють запасам, а суми за стовпцями дорівнюють заявкам.

Отриманий план є опорним, тому що число ненульових перевезень дорівнює  $m + n - 1 = 3 + 5 - 1 = 7$ , а число нульових перевезень дорівнює  $(n-1) \cdot (m-1) = (5-1) \cdot (3-1) = 8$ .

План можна покращити (табл. 6.2), якщо зменшити перевезення в дорогій клітці, наприклад (1.1), і збільшити в дешевій (3.1).

Щоб при цьому план залишався опорним, необхідно одну з вільних кліток зробити базисною, а одну з базисних - вільною.

### 6.3 Метод потенціалів

Одним з методів розв'язання транспортної задачі є метод, що одержав назву **методу потенціалів**. Уперше він був запропонований в 1949 р. Л.В. Канторовичем і М.К. Гавуріним. Пізніше на базі загальних ідей лінійного програмування аналогічний метод був запропонований Дж. Данцигом.

Точно так само як транспортна задача є окремим випадком задачі ЛП, так і метод потенціалів, загалом кажучи, може трактуватися як різновид симплексних процедур. Він являє собою ітеративний процес, на кожному кроці якого розглядається деякий поточний базисний план, перевіряється його оптимальність, і при необхідності здійснюється, перехід до «кращого» базисного плану.

Алгоритм методу потенціалів починається з вибору деякого припустимого базисного плану. Якщо початковий опорний план має  $m+n-1$  додатних перевезень, то він називається **невиродженим**. Якщо ж опорний план має менше за  $m+n-1$  додатних перевезень, то він називається **виродженим**.

У цьому початковому опорному плані кожному пункту ставлять у відповідність деяке число, назване його **попереднім потенціалом**. Якщо даний план не вироджений (число ненульових базисних кліток дорівнює  $m+n-1$ ), то по ньому можна так визначити потенціали  $u_i$  і  $v_j$ , щоб для кожної базисної клітки (тобто для тої, у якій  $x_{ij} > 0$ ) виконувалася умова

$$v_j - u_i = c_{ij}, \quad \text{якщо } x_{ij} > 0. \quad (6.4)$$

**Теорема 6.1.** Якщо план ТЗ є оптимальним, то йому відповідає система з  $m+n$  чисел, що задовольняють умовам:

$$\begin{aligned} u_i + v_j &= c_{ij} & \text{для } x_{ij} > 0, \\ u_i + v_j &\leq c_{ij} & \text{для } x_{ij} = 0, \quad i = \overline{1, m}; \quad j = \overline{1, n}. \end{aligned}$$

де  $u_i$  і  $v_j$  - потенціали постачальників і споживачів відповідно.

Таким чином, якщо хоча б для однієї вільної клітки

$$\Delta_{ij} = u_i + v_j - c_{ij} > 0, \quad (6.5)$$

то план не оптимальний і потребує поліпшення.

Оскільки система (6.4) містить  $m+n-1$  рівнянь й  $m+n$  невідомих, то один з потенціалів можна задати довільно (наприклад, дорівняти  $v_j$  або  $u_i$ , до нуля). Після цього інші невідомі  $u_i$  і  $v_j$  - визначаються однозначно.

Розглянемо процес визначення потенціалів поточного плану транспортної задачі на прикладі. У таблиці 5.3 переписані умови задачі з таблиці 6.1 і її припустимий базисний план, побудований за методом північно-західного кута з таблиці 6.2.

Потенціал першого пункту виробництва приймаємо рівним нулю ( $u_1=0$ ). Тепер, знаючи його, ми можемо визначити потенціали для всіх пунктів призначення, пов'язаних з першим пунктом виробництва ненульовими перевезеннями. У цьому випадку їх три (це перший, другий і третій пункти), отримуємо:

$$v_1 = c_{11} - u_1 = 7 - 0 = 7; \quad v_2 = c_{12} - u_1 = 5 - 0 = 5; \quad v_3 = c_{13} - u_1 = 2 - 0 = 2.$$

Маючи  $v_3$  і з огляду на те, що в другому рядку таблиці існують ненульові компоненти  $x_{23}$  і  $x_{24}$ , можна визначити  $u_2 = c_{23} - v_3 = 4 - 2 = 2$ ,  $v_4 = c_{24} - u_2 = 6 - 2 = 4$ , після чого з'являється можливість розрахувати  $u_3 = c_{34} - v_4 = 2 - 4 = -2$  і, нарешті,  $v_5 = c_{35} - u_3 = 3 - (-2) = 5$ . У результаті отримуємо повну систему потенціалів, показану в таблиці 6.5.

Таблиця 6.5 - Визначення потенціалів для початкового опорного плану

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=5$	
		<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>B<sub>5</sub></b>	<b>Запаси</b>
$u_1=0$	<b>A<sub>1</sub></b>	7 30	5 50	2 45	8	7	125
$u_2=2$	<b>A<sub>2</sub></b>	8	9	4 55	6 5	9	60
$u_3=-2$	<b>A<sub>3</sub></b>	5	1	9	2 35	3 80	115
	<b>Заявки</b>	30	50	100	40	80	300

Для вільних кліток транспортної таблиці обчислюють величини  $\Delta_{ij} = v_j + u_i - c_{ij}$ . У таблиці 6.6 вони виписані для всіх небазисних кліток під цінами.

Таблиця 6.6 - Перевірка оптимальності поточного плану (обчислення оцінок  $\Delta_{ij}$ )

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=5$	
		<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>B<sub>5</sub></b>	<b>Запаси</b>
$u_1=0$	<b>A<sub>1</sub></b>	7 30	5 50	2 45	8 0	7 -2	125
$u_2=2$	<b>A<sub>2</sub></b>	8 1	9 -2	4 55	6 5	9 -2	60
$u_3=-2$	<b>A<sub>3</sub></b>	5 0	1 2	9 -9	2 35	3 80	115
	<b>Заявки</b>	30	50	100	40	80	300

Відповідно до теореми 6.1, якщо всі  $\Delta_{ij} \leq 0$ , то план оптимальний, у протилежному випадку, якщо існує хоча б одна клітка, для якої  $\Delta_{ij} > 0$ , то він може бути поліпшений. Процес «поліпшення» плану полягає у визначенні клітки, що вводиться, і клітки, що виводиться. У цьому простежується змістовна аналогія методу з відповідними пунктами симплекс-процедур.

Кандидатом на введення може бути будь-яка клітка, у якій  $\Delta_{ij} > 0$ , оскільки після введення її до базису буде забезпечена рівність  $v_j + u_i = c_{ij}$ . Для визначеності рекомендується брати ту клітку, у якій оцінка  $\Delta_{ij}$  максимальна. У розглянутому нами прикладі це буде клітка (3, 2).

Виведена клітка визначається за допомогою так званого ланцюжка перетворення плану, що описує характер перерозподілу вантажних потоків. У відповідності із властивостями транспортної задачі для невивродженого базисного плану в поточній таблиці можна утворити замкнутий ланцюжок, що складається з вертикальних і горизонтальних ланок, однієї з вершин якої є обрана вільна клітка, а інші - зайняті клітки. У таблиці 6.7 показаний ланцюжок перетворення поточного плану щодо клітки, яка вводиться до нього, (3, 2).



Таблиця 6.7 - Перетворення поточного плану

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=5$	
		<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>B<sub>5</sub></b>	<b>Запаси</b>
$u_1=0$	<b>A<sub>1</sub></b>	7 30	5 50 -Θ	2 45 +Θ	8 4	7 5	125
$u_2=2$	<b>A<sub>2</sub></b>	8 5	9 3	4 55 -Θ	6 5 +Θ	9 3	60
$u_3=-2$	<b>A<sub>3</sub></b>	5 9	1 +Θ 7	9 4	2 35 -Θ	3 80	115
	<b>Заявки</b>	30	50	100	40	80	300

Логіка алгоритму побудови ланцюжка досить проста: «вийшовши» із клітки (3, 2) у горизонтальному напрямку, ми повинні «зупинитися» у тій зайнятій клітці плану, з якої зможемо рухатися далі за вертикаллю. У даному прикладі цій вимозі задовольняють як клітка (3, 4), так і клітка (3, 5). Однак ланцюжок від (3, 5) не може бути продовжений далі, у той час як рухаючись від (3, 4) за вертикаллю до (2, 4) і далі до (2, 3), ми повертаємося через клітки (1, 3) і (1, 2) до вихідної клітки (3, 2) і утворимо замкнений цикл.

У побудованому ланцюжку, починаючи із клітки, що вводиться (яка вважається першою), позначаються вершини: непарні - «+Θ», а парні «-Θ». Знаком «+» відзначаються ті клітки, у яких обсяги перевезень повинні збільшитися (такою, зокрема, є клітка, що вводиться до плану, оскільки вона повинна стати базисною). Знаком «-» - ті клітки, у яких перевезення зменшуються з метою збереження балансу. Серед множини кліток, позначених знаком «-», вибирається клітка з найменшим значенням  $x_{ij}$ . Вона і стає кандидатом на вивід, тому що зменшення обсягу перевезень на більшу величину може привести до від'ємних значень  $x_{ij}$  в інших «мінусових» клітках. Потім провадиться перерахування плану за ланцюжком: до обсягів перевезень у клітках, позначених знаком «+», додається обсяг  $\Theta$ , а з обсягів кліток, позначених знаком «-», він віднімається. У результаті введення однієї клітки й виводу іншої утворюється новий базисний план, для якого на наступній ітерації описані вище дії повторюються.

У нашому прикладі знаком «-» відзначені клітки (3, 4), (2, 3) і (1, 2), причому  $x_{34}=35$ ,  $x_{23}=55$ ,  $x_{12}=50$ . Обчисливши значення  $\Theta = \min\{x_{34}, x_{23}, x_{12}\} = 35$ , здійснюємо перетворення і переходимо до наступного базисного плану, показаного в таблиці 6.8.

Таблиця 6.8 - Оцінка оптимальності наступного базисного плану

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=7$	
		<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>B<sub>5</sub></b>	<b>Запаси</b>
$u_1=0$	<b>A<sub>1</sub></b>	7 30	5 15	2 80	8 -4	7 0	125
$u_2=2$	<b>A<sub>2</sub></b>	8 1	9 -2	4 20	6 40	9 0	60
$u_3=-4$	<b>A<sub>3</sub></b>	5 -2	1 35	9 -11	2 -2	3 80	115
	<b>Заявки</b>	30	50	100	40	80	300

Для знов отриманого плану повторюються дії стандартної ітерації: розраховуються потенціали й оцінки для небазисних кліток транспортної таблиці. Як можна бачити, план у таблиці 5.6 також не є оптимальним (у клітці (2, 1)  $\Delta_{21} = 1 > 0$ ), тому знову будуємо ланцюжок перетворення плану й переходимо до наступного базисного плану за ланцюжком в таблиці 6.9.

Таблиця 6.9 - Перетворення поточного базисного плану

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=7$	
		<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>B<sub>5</sub></b>	<b>Запаси</b>
$u_1=0$	<b>A<sub>1</sub></b>	7 30; - $\Theta$	5 15	2 80; + $\Theta$	8 -4	7 0	125
$u_2=2$	<b>A<sub>2</sub></b>	8 + $\Theta$ 1	9 -2	4 20; - $\Theta$	6 40	9 0	60
$u_3=-4$	<b>A<sub>3</sub></b>	5 -2	1 35	9 -11	2 -2	3 80	115
	<b>Заявки</b>	30	50	100	40	80	300

Визначивши  $\Theta=20$ , одержимо

Таблиця 6.10 - Оптимальний базисний план

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=5$	$v_5=7$	
		<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>B<sub>5</sub></b>	<b>Запаси</b>
$u_1=0$	<b>A<sub>1</sub></b>	7 10	5 15	2 100	8 -3	7 0	125
$u_2=1$	<b>A<sub>2</sub></b>	8 20	9 -3	4 -1	6 40	9 -1	60
$u_3=-4$	<b>A<sub>3</sub></b>	5 -2	1 35	9 -11	2 -1	3 80	115
	<b>Заявки</b>	30	50	100	40	80	300

З транспортної таблиці 6.10 видно, що отриманий план оптимальний, тому що всі оцінки для небазисних кліток  $\Delta_{ij} \leq 0$ , тобто  $u_i + v_j$  не перевищують відповідних цін  $c_{ij}$ . За даним планом обчислюється оптимальне (найменше) значення сумарних витрат на перевезення

$$L^* = 10x_7 + 15x_5 + 100x_2 + 20x_8 + 40x_6 + 35x_1 + 80x_3 = 1020.$$

## 6.4 Випадок виродження

Зупинимося на ситуації виникнення виродженого плану. Якщо задача вироджена, то на якомусь етапі розв'язання може виявитися, що таблиця містить менш  $m+n-1$  заповнених кліток. Це, зокрема, може відбутися, якщо однакове мінімальне значення буде досягнуто відразу на декількох клітках, позначених знаком «-». Для подолання виродженості в транспортній задачі поточний план доповнюють необхідною кількістю нульових кліток (фіктивними перевезеннями) таким чином, щоб вони дозволяли розрахувати повну систему потенціалів, і далі діяти відповідно до правил описаного вище алгоритму. Тобто, якщо не вистачає  $k$  заповнених кліток, то їх вважають фіктивно заповненими. Фактично такий прийом є аналогом методу збурювань

для транспортної задачі як окремого випадку ЗЛП. До такого висновку легко прийти, якщо покласти, що фіктивні клітки, які додаються, містять деякий малий обсяг  $\varepsilon$ .

Розглянемо приклад. З трьох кар'єрів до чотирьох склозаводів возять глину. Вартості перевезень, потужності кар'єрів і потреби заводів наведені в таблиці 6.11.

Таблиця 6.11 - Вихідні дані

	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>Запаси</b>
<b>A<sub>1</sub></b>	3	9	7	4	50
<b>A<sub>2</sub></b>	6	8	10	6	65
<b>A<sub>3</sub></b>	5	4	7	6	50
<b>Потреби</b>	50	20	65	30	165

Опорний план визначимо за методом найменшої вартості.

Таблиця 6.12 - Початковий опорний план

		$v_1=3$	$v_2=5$	$v_3=8$	$v_4=4$	
		<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>Запаси</b>
$u_1=0$	<b>A<sub>1</sub></b>	3 50	9 -4	$+ \theta$ 7 1	$- \theta$ 4 [0]	50
$u_2=2$	<b>A<sub>2</sub></b>	6 -1	8 -1	$- \theta$ 10 35	$+ \theta$ 6 30	65
$u_3=-1$	<b>A<sub>3</sub></b>	5 -3	4 20	7 30	6 -3	50
	<b>Потреби</b>	50	20	65	30	165

У таблиці 6.12 заповнених кліток 5, а необхідно  $m+n-1 = 3+4-1 = 6$ . Потрібна одна фіктивно заповнена клітка. Будемо вважати клітку (1, 4) заповненою. Перевіримо план на оптимальність.

План не оптимальний, тому що  $\Delta_{13} = 1$ .  $\theta = \min(35, 0) = 0$ . План залишається таким же, але фіктивно заповненою буде клітка (1, 3). Перевіримо його на оптимальність.

Таблиця 6.13 - Перевірка базисного плану на оптимальність

		$v_1=3$	$v_2=4$	$v_3=7$	$v_4=3$	
		<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>B<sub>4</sub></b>	<b>Запаси</b>
$u_1=0$	<b>A<sub>1</sub></b>	3 50	9 -5	7 0	4 -1	50
$u_2=3$	<b>A<sub>2</sub></b>	6 0	8 -1	10 35	6 30	65
$u_3=0$	<b>A<sub>3</sub></b>	5 -2	4 20	7 30	6 -3	50
	<b>Потреби</b>	50	20	65	30	165

Всі  $\Delta_{ij} \leq 0$ , отже план є оптимальним. Вартість перевезень при такому плані становить  $L = 50 \cdot 3 + 20 \cdot 4 + 0 \cdot 7 + 35 \cdot 10 + 30 \cdot 7 + 30 \cdot 6 = 970$ .

## 6.5 Транспортна задача за критерієм часу

Транспортна задача, крім використання її з метою мінімізації витрат при транспортуванні вантажів, застосовується для оптимізації цілого ряду економічних процесів.

Природним застосуванням транспортної задачі є задача мінімізації часу транспортування. Наприклад, на підприємстві є задана кількість виробничих ліній, що працюють із заданою продуктивністю. Є також задана кількість складів продукції. Відомий час транспортування від кожної лінії до кожного складу. Задача зводиться до вибору таких маршрутів, при використанні яких час, затрачений на транспортування, буде мінімальним.

Іншою задачею є вибір оптимального варіанта з використання наявного встаткування. Задача полягає у виборі оптимального плану виробництва виробів при мінімізації часу, необхідного на їхнє виготовлення. Ця задача може бути сформульована в такий спосіб. Є задана кількість видів устаткування, на кожному з яких може бути виготовлена деяка задана кількість виробів. Відома продуктивність для кожного типу устаткування й кожного виробу. Необхідно скласти план, при якому підприємство затратить мінімальну кількість часу на виготовлення всіх виробів.

Розглянемо приклад. Визначити мінімальний час виготовлення трьох видів виробів на трьох типах устаткування. Продуктивність устаткування ( $a_i$ ) і план виробництва виробів ( $b_i$ ) наведені в таблиці 6.14.

Таблиця 6.14 - Вихідні дані

	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>План</b>
<b>A<sub>1</sub></b>	15	7	8	240
<b>A<sub>2</sub></b>	9	4	11	80
<b>A<sub>3</sub></b>	6	3	7	180
<b>Продуктивність</b>	200	160	140	500

Задача є збалансованою. Знайдемо початковий базисний план, використовуючи метод мінімальної вартості. Найменша продуктивність, що дорівнює 3, записана в клітці (3, 2). Потужність устаткування третього типу дорівнює 180, а необхідна кількість виробів другого типу дорівнює 160, тобто  $\min\{180, 160\} = 160$ . Тому в клітку (3, 2) уписуємо число 160. Другий стовпець випадає з подальшого розгляду, тому що продуктивність устаткування другого типу вичерпана. Аналогічно заповнюємо інші клітки таблиці 6.15.

Таблиця 6.15 - Визначення початкового опорного плану

		$v_1=15$	$v_2=12$	$v_3=8$	
		<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>План</b>
$u_1=0$	<b>A<sub>1</sub></b>	15 -Θ 100	7 +Θ 5	8 140	240
$u_2=-6$	<b>A<sub>2</sub></b>	9 80	4 2	11 -9	80
$u_3=-9$	<b>A<sub>3</sub></b>	6 +Θ 20	3 -Θ 160	7 -8	180
	<b>Продуктивність</b>	200	160	140	500

Відповідно до отриманого плану час виготовлення всіх виробів становить  $L = 100 \cdot 15 + 140 \cdot 8 + 80 \cdot 9 + 20 \cdot 6 + 160 \cdot 3 = 3940$ . Визначимо потенціали  $u_i$  і  $v_j$  для зайнятих кліток:  $u_1=0$ ;  $v_1=15$ ;  $v_3=8$ ;  $u_2=9-15=-6$ ;  $u_3=6-15=-9$ ;  $v_2=3-(-9)=12$ . Обчислимо оцінки для вільних кліток:  $\Delta_{12}=(12+0)-7=5$ ;  $\Delta_{22}=(12-6)-4=2$ ;  $\Delta_{23}=(-6+8)-11=-9$ ;  $\Delta_{33}=(-9+8)-7=-8$ .

Оскільки оцінки вільних кліток (1, 2) і (2, 2) додатні, план не оптимальний. Перейдемо до нового базисного плану, побудувавши цикл для клітки (1, 2),  $\Theta = \min\{100, 160\}=100$ . Новий базисний план наведений у таблиці 6.16. Із клітки (1,1) вироби в кількості 100 од. переносимо до клітки (1,2). Із клітки (3,2) вироби в кількості 100 од. переносимо до клітки (3,1).

Таблиця 6.16 - Визначення наступного базисного плану

		$v_1=10$	$v_2=7$	$v_3=8$	
		<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>План</b>
$u_1=0$	<b>A<sub>1</sub></b>	15 -5	7 100	8 140	240
$u_2=-1$	<b>A<sub>2</sub></b>	- $\Theta$ 80	-9 4 2	- $+\Theta$ 11 -4	80
$u_3=-4$	<b>A<sub>3</sub></b>	+ $\Theta$ 120	6 60	-3 7	180
	<b>Продуктивність</b>	200	160	140	500

Відповідно до отриманого плану час виготовлення всіх виробів складе  $L=100 \cdot 7 + 140 \cdot 8 + 80 \cdot 9 + 120 \cdot 6 + 60 \cdot 3 = 3440$ . Перевірка плану на оптимальність показує, що він не оптимальний, тому що в клітці (2,2)  $\Delta_{22}$  додатна. Побудуємо цикл для цієї клітки й перетворимо план при  $\Theta=60$ . Новий базисний план наведений у таблиці 6.17.

Таблиця 6.17 - Визначення наступного базисного плану

		$v_1=12$	$v_2=7$	$v_3=8$	
		<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>	<b>B<sub>3</sub></b>	<b>План</b>
$u_1=0$	<b>A<sub>1</sub></b>	15 -3	7 100	8 140	240
$u_2=-3$	<b>A<sub>2</sub></b>	9 20	4 60	11 -6	80
$u_3=-6$	<b>A<sub>3</sub></b>	6 180	3 -2	7 -5	180
	<b>Продуктивність</b>	200	160	140	500

Перевірка показує оптимальність плану. Відповідно до отриманого плану час виготовлення всіх виробів складе  $L = 100 \cdot 7 + 140 \cdot 8 + 20 \cdot 9 + 60 \cdot 4 + 180 \cdot 6 = 3320$ , і він є мінімальним.

### Контрольні запитання

1. Які специфічні властивості дозволяють виділити транспортні задачі в окремий клас з множини задач лінійного програмування?
2. Опишіть методи побудови припустимого плану транспортної задачі.
3. Скільки ненульових елементів повинен містити не вироджений базисний план транспортної задачі?

4. Сформулюйте критерій оптимальності для припустимого плану транспортної задачі.
5. На чому заснований метод потенціалів?
6. З чого впливає критерій оптимальності припустимого плану транспортної задачі?
7. Перелічіть основні етапи методу потенціалів.
8. Які умови повинні бути дотримані при побудові ланцюжка перетворення плану в методі потенціалів?
9. Що треба робити при виникненні ситуації виродженості поточного плану в транспортній задачі?

## ТЕМА 7

### ЦІЛОЧИСЛОВІ ЗАДАЧІ ЛІНІЙНОГО ПРОГРАМУВАННЯ. ОСНОВНІ МЕТОДИ ЇХ РОЗВ'ЯЗАННЯ І АНАЛІЗУ

#### 7.1 Типи задач дискретного програмування

Цілочислове лінійне програмування орієнтоване на розв'язання задач лінійного програмування, у яких всі або деякі змінні повинні приймати цілочислові (або дискретні) значення. Багато економічних задач характеризуються тим, що обсяги керованих ресурсів можуть приймати тільки цілі значення. До цілочислового програмування належать також задачі, у яких змінні можуть приймати тільки два значення – 0 або 1 (булеві або бінарні змінні). До задач цілочислового програмування належать задачі про призначення, про найкоротший шлях і інші. У загальному вигляді задача цілочислового програмування може бути сформульована як задача знаходження максимуму (або мінімуму) цільової функції  $L(x_1, x_2, \dots, x_n)$  на множині  $D$ , зумовленій системою обмежень. Загальна задача цілочислового програмування формулюється в такий спосіб.

Знайти найбільше або найменше значення лінійної функції

$$L = \sum_{j=1}^n c_j x_j \rightarrow \text{extr} \quad (7.1)$$

за умови

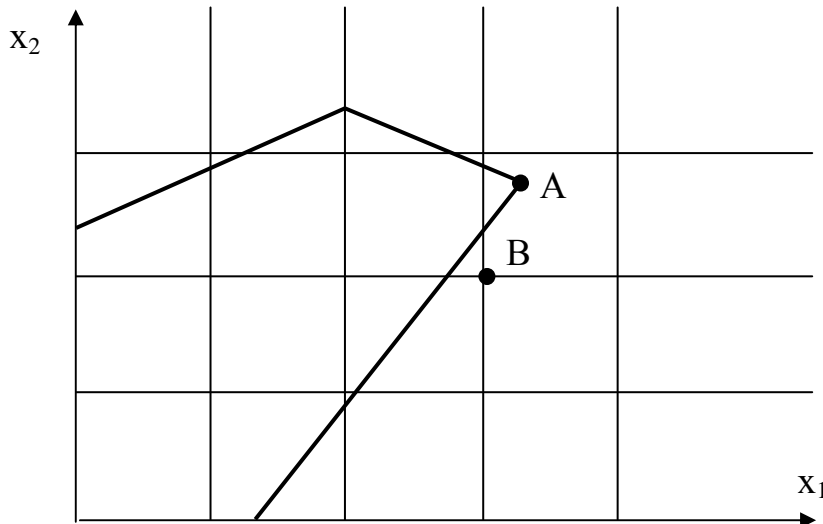
$$\sum_{j=1}^n a_{ij} x_j = b_i; \quad i = \overline{1, m} \quad (7.2)$$

$$x_j \geq 0, \quad x_j - \text{целые}, \quad j = \overline{1, n}.$$

Умова  $x_j$  – цілі називається умовою дискретності. Особливе місце серед дискретних задач займає цілочислова задача лінійного програмування в канонічній формі (ЦКЗЛП). У деяких ситуаціях вимога «цілочислове» може бути накладена лише на деякі змінні  $x_j$ , що кардинально не змінює характеру задачі.

Принципова складність, викликувана наявністю умов цілочисловості в системі обмежень оптимізаційної задачі, полягає в тому, що в значній кількості випадків неможливо замінити дискретну задачу її безперервним аналогом і,

знайшовши відповідний розв'язок, округлити його компоненти до найближчих цілих значень. На рисунку 7.1 показано, що при округленні оптимального плану  $x^*$  звичайної задачі ЛП до цілих значень виходить точка  $B$ , яка не належить області припустимих планів. Якщо навіть оптимальний план безперервної задачі, округлений до цілих значень компонент, виявиться припустимим, то цільова функція може поводитися так, що її значення буде на ньому істотно «гірше», ніж на оптимальному плані цілочислової задачі.



Рисунку 7.1 – Розв'язок цілочислової задачі

Перелічені проблеми визначили необхідність розробки спеціальних методів розв'язання дискретних і цілочислових задач.

У літературі, як правило, виділяють наступні класи дискретних оптимізаційних задач:

- задача з неділимостями;
- екстремальні комбінаторні задачі;
- задачі з розривними цільовими функціями;
- задачі на незв'язних і неопуклих областях та ін.

**Задача з неділимостями.** У переважній більшості випадків наявність умов неділимості визначається фізичними властивостями об'єктів, наприклад, вони можуть з'явитися як додаткові обмеження в задачі виробничого планування, якщо в ній здійснюється управління випуском крупної штучної продукції. Класичним представником задач даного класу є так звана задача про ранець. Вона полягає в тому, що солдат (або турист), що збирається в похід, може нести вантаж вагою, яка не перевищує  $W$  кг. Цей вантаж може складатися з набору предметів  $n$  типів, кожний предмет  $j$ -го типу важить  $w$  кг і характеризується деякою «корисністю». Треба визначити, скільки предметів кожного виду потрібно покласти в ранець, щоб його сумарна корисність була максимальною. До такого формулювання можуть бути зведені багато економічних задач. У літературі ця задача також відома як задача про завантаження судна.

**Комбінаторні задачі.** До даного класу належать задачі оптимізації функції, заданої на кінцевій множині, елементами якої служать вибірки з  $n$

об'єктів. Класичним представником такої задачі є задача про комівояжера. Вона полягає в складанні маршруту відвідування торговельним агентом, що перебуває в деякому початковому пункті,  $n$  інших міст за умови, що задано матрицю вартостей переїздів з міста до міста. Причому припустимим є маршрут, який передбачає однократне відвідування всіх міст і повернення у вихідний пункт. Найкращий маршрут повинен мінімізувати сумарну вартість переїздів. Кожний припустимий маршрут можна ототожнити з перестановкою  $n$  чисел. Задача комівояжера має велику кількість змістовних аналогів. Зокрема, задача розробки графіка переналагодження встаткування, що може випускати різні типи виробів, але вимагає певних витрат (часових або матеріальних) при переході з одного технологічного режиму на іншій.

**Задача з розривними цільовими функціями.** Багато економічних систем характеризуються наявністю так званих постійних витрат, які повинні бути зроблені незалежно від обсягу виробництва. Урахування у моделях цих і подібних факторів приводить до появи в них цільових функцій, які не є безперервними. Як приклад може бути наведено транспортну задачу з фіксованими доплатами. У цьому випадку цільова функція сумарних витрат на перевезення містить «стрибкоподібні» розриви, що істотно утрудняє її мінімізацію.

Методи розв'язання задач цілочисельного лінійного програмування засновані на використанні обчислювальних можливостей методів лінійного програмування. Звичайно алгоритми цілочислового програмування включають три кроки:

1. «Ослаблення» простору припустимих розв'язків задачі шляхом відкидання вимоги цілочисловості. У результаті виходить звичайна задача лінійного програмування.

2. Визначення оптимального розв'язку задачі лінійного програмування.

3. Маючи отриманий оптимальний розв'язок, додають спеціальні обмеження, які ітераційним шляхом змінюють простір припустимих розв'язків задачі лінійного програмування таким чином, щоб в остаточному підсумку вийшов оптимальний розв'язок, що задовольняє вимогам цілочисельності.

Розроблено два методи побудови спеціальних обмежень:

- метод Гоморі (або метод площин, що відтинають, він був уперше запропонованим Р.Гоморі у 1957-1958 р.);
- метод віток і границь.

Помітимо також, що досить ефективний і широко застосовуваний підхід до розв'язання цілочислових задач заснований на зведенні їх до задач транспортного типу. Це пояснюється тим, що якщо в умовах транспортної задачі значення запасів і потреб є цілочисловими, то цілочисловим буде й оптимальний план.



## 7.2 Метод Гоморі

Сутність методу Гоморі полягає в тому, що спочатку задача вирішується без урахування умови цілочисловості. Якщо отриманий у результаті оптимальний план  $x^*$  містить тільки цілі компоненти, задачу вирішено. У протилежному випадку до системи обмежень задачі додають нове обмеження, що володіє наступними властивостями:

- воно повинне бути лінійним;
- повинне відтинати знайдений цілочисловий оптимальний план  $x^*$ ;
- не повинне відтинати жодного цілочислового плану.

Додаткове обмеження, що володіє зазначеними властивостями, називається **правильним відсіканням**. Далі задача вирішується з урахуванням нового обмеження. Після цього, якщо буде потреба, додається ще одне обмеження і т.д.

Геометрично додавання кожного лінійного обмеження відповідає проведенню прямої (гіперплощини), що відтинає від багатокутника (багатогранника) розв'язків деяку його частину разом з оптимальною точкою з нецілими координатами, але не зачіпає ні однієї із цілих точок цього багатогранника. У результаті новий багатогранник розв'язків містить всі цілі точки, утримувані в первісному багатограннику розв'язків, і відповідно отриманий при цьому багатограннику оптимальний розв'язок буде цілочисловим.

Один з алгоритмів розв'язання задачі лінійного цілочислового програмування, запропонований Гоморі, заснований на симплексному методі й використовує досить простий спосіб побудови правильного відсікання.

Нехай задача лінійного програмування має кінцевий оптимум і на останньому кроці її розв'язання симплексним методом отримані наступні рівняння, у яких базисні змінні  $x_1, x_2, \dots, x_m$  виражені через вільні змінні  $x_{m+1}, x_{m+2}, \dots, x_n$  оптимального розв'язку

[illegible]

де оптимальним розв'язком задачі є  $x^* = (b_1, b_2, \dots, b_m, 0, 0, \dots, 0)$ , у якому, наприклад,  $b_i$  – нецілий компонент.

Нагадаємо, що цілою частиною числа  $b$  називається найбільше ціле число  $[b]$ , ще не перевищує  $b$ , а дробовою частиною числа  $b$  - число  $\{b\}$ , що дорівнює різниці між цим числом і його цілою частиною, тобто  $\{b\} = b - [b]$ .

Сформувавши за  $i$ -м рівнянням системи (6.3) додаткове обмеження

$$\{b_i\} - \{a_{i+m+1}\} \{x_{m+1}\} - \dots - \{a_{i+n}\} \{x_n\} \leq 0, \quad (7.4)$$

МОЖНА ДОВЕСТИ, ЩО ВОНО МАЄ ВСІ ВЛАСТИВОСТІ ПРАВИЛЬНОГО ВІДСІКАННЯ.

Для розв'язання задачі цілочислового програмування методом Гоморі використовується наступний алгоритм.

1. Вирішити задачу симплексним методом без урахування умови цілочисловості. Якщо всі компоненти оптимального плану цілі, то він є оптимальним і для задачі цілочислового програмування. Якщо перша задача є нерозв'язною (тобто не має кінцевого оптимуму або умови її суперечливі), то і друга задача також нерозв'язна.

2. Якщо серед компонентів оптимального розв'язку є нецілі, то вибрати компоненту з найбільшою цілою частиною і за відповідним їй рівнянням системи (7.3) сформулювати правильне відсікання (7.4).

3. Отриману нерівність перетворити на рівносильне рівняння шляхом введення додаткової невід'ємної цілочислової змінної

$$\{b_i\} - \{a_{i_{m+1}}\}x_{m+1} - \dots - \{a_{i_n}\}x_n + x_{n+1} \quad (7.5)$$

і включити його до системи обмежень. Необхідно пам'ятати, що після включення до системи обмежень додаткового рівняння, яке відповідає правильному відсіканню, завжди буде виходити неприпустимий базисний розв'язок. Для одержання припустимого базисного розв'язку потрібно перевести в базисні змінні одну з вільних змінних.

4. Отриману розширену задачу вирішити симплексним методом. Якщо знайдений оптимальний план буде цілочисловим, то задача цілочислового програмування вирішена. У протилежному випадку треба повернутися до п.2 алгоритму.

Якщо задачу розв'язано в цілих числах, то після кінцевого числа кроків оптимальний цілочисловий план буде знайдений.

Якщо в процесі розв'язання з'явиться рівняння (що виражає базисну змінну через вільні) з нецілим вільним членом і цілими іншими коефіцієнтами, то відповідне рівняння не має розв'язання в цілих числах. У цьому випадку й задача не має цілочислового оптимального розв'язку.

Розглянемо особливості застосування методу Гоморі на прикладі. Нехай є задача з наступними умовами:

$$\begin{aligned} L &= 3x_1 + x_2 + 2x_3 \rightarrow \max, \\ x_1 + 2x_2 + 4x_3 &= 9, \\ 3x_1 + x_2 + 5x_3 &= 10, \\ x_j &\geq 0, \quad x_j - \text{цілі}, \quad j = \overline{1,3}. \end{aligned}$$

Використовуючи звичайний симплекс-алгоритм, вирішуємо безперервний аналог вихідної задачі, у якій ігноруються умови цілочисловості. Як вихідний базис можна взяти перший і другий стовпці. На його основі заповнюється симплекс-таблиця 7.1.

Таблиця 7.1 – Безперервна модель, що відповідає цілочисловій задачі

Базис	$C_{\text{баз}}$	$C_j$	3	1	2
		$B$	$A_1$	$A_2$	$A_3$
$A_1$	3	11/5	1	0	6/5
$A_2$	1	17/5	0	1	7/5
$L_j$		10	3	1	5
$\Delta_j$			0	0	3

Як видно з рядка оцінок, даний базис є оптимальним, однак відповідний йому план  $x^* = \left(\frac{11}{5}; \frac{17}{5}; 0\right)$  не є цілочисловим, тому вибираємо рядок, що містить перший нецілий елемент, і відповідно до формули (7.4) будуємо обмеження, що відтинає:

$$\left\{\frac{11}{5}\right\} - \left\{\frac{6}{5}\right\}x_3 \leq 0.$$

Увівши додаткову цілочислову змінну  $x_4 \geq 0$ , одержимо рівносильне нерівності рівняння

$$-\frac{1}{5}x_3 + x_4 = -\frac{1}{5},$$

яке необхідно включити до системи обмежень. Уведемо його до симплекс-таблиці 7.2.

Таблиця 7.2 – Нова система обмежень цілочислової задачі

Базис	$C_{баз}$	$C_i$	3	1	2	0
		$B$	$A_1$	$A_2$	$A_3$	$A_4$
$A_1$	3	11/5	1	0	6/5	0
$A_2$	1	17/5	0	1	7/5	0
$A_4$	0	-1/5	0	0	-1/5	1
$L_j$		10	3	1	5	0
$\Delta_i$			0	0	3	0

Врахуємо, що отриманий базисний розв'язок є неприпустимим (компонента  $b_3$  від'ємна). Для одержання припустимого базисного розв'язку переведемо в базисні змінні вільну змінну  $x_3$ .

Базис	$C_{баз}$	$C_i$	3	1	2	0
		$B$	$A_1$	$A_2$	$A_3$	$A_4$
$A_1$	3	1	1	0	0	6
$A_2$	1	2	0	1	0	7
$A_3$	2	1	0	0	1	-5
$L_j$		7	3	1	2	15
$\Delta_i$			0	0	0	15

Отриманий план  $x^* = (1, 2, 1, 0)$  є не тільки оптимальним (всі  $b_i > 0$ ), але й повністю складається з цілочислових компонент, тобто розв'язок задачі знайдено,  $\max L = 7$ .

### 7.3 Метод віток і границь

Метод віток і границь - один з комбінаторних методів. Його суть полягає в упорядкованому переборі варіантів і розгляданні тих з них, які за певними ознаками є перспективними. Уперше даний метод запропонували в 1960 р. Ленг і Дойг, а його «друге народження» відбулося в 1963 р. у зв'язку з виходом роботи Літтла, Мурті, Суїні й Керел, присвяченої розв'язанню задачі про комівояжера.

Метод віток і границь полягає в тому, що множина припустимих розв'язків (планів) деяким способом розбивається на підмножини, кожна з яких цим же способом знову розбивається на підмножини. Процес триває доти, поки не буде отриманий оптимальний цілочисловий розв'язок вихідної задачі.

Якщо задача максимізації лінійної функції  $L$  вирішена симплексним методом без урахування цілочисловості змінних, то стають відомими нижня й верхня границя для кожної цілочислової змінної  $x_j$ :  $[x_j] < x_j < [x_j] + 1$  і нижня границя лінійної функції  $L_0$ , тобто при будь-якому плані  $x$   $L(x) \geq L_0$ . Тоді з області припустимих значень, наприклад, змінної  $x_r$  виключають область  $[x_r] < x_r < [x_r] + 1$ , де  $[x_r]$  – ціла частина числа  $x_r$ . У результаті це дозволяє сформулювати дві задачі, що відрізняються тим, що до одної з них додано обмеження  $x_r \leq [x_r]$ , а до іншої – обмеження  $x_r \geq [x_r] + 1$ .

Сформовані задачі вирішують у будь-якому порядку. Залежно від отриманого розв'язку список таких задач може розширюватися або зменшуватися. Якщо в ході розв'язання будь-якої із задач отриманий нецілочисловий й оптимальний план, для якого  $L(x) \leq L_0$ , то дану задачу виключають зі списку. Якщо  $L(x) \geq L_0$ , то з даної задачі формують дві нові задачі.

Розглянемо приклад:

$$\begin{aligned} L &= 5x_1 + 4x_2 \rightarrow \max, \\ x_1 + x_2 &\leq 5, \\ 10x_1 + 6x_2 &\leq 45, \\ x_j &\geq 0, \quad x_j - \text{цілі}, \quad j = \overline{1,2}. \end{aligned}$$

Вирішимо задачу шляхом відкидання умов цілочисловості. Використаємо функцію Excel «Пошук рішення» і одержимо

$$x_1=3,75; x_2=1,25; L=23,75.$$

Виберемо одну з цілочислових змінних, значення якої в оптимальному розв'язку не є цілочисловим, наприклад,  $x_1$ . Очевидно, що область  $3 < x_1 < 4$  простору припустимих розв'язків не містить цілочислових значень змінної  $x_1$ , і отже може бути виключена з розгляду. Це еквівалентно заміні вихідної задачі двома новими задачами. В одну з них додамо обмеження  $x_1 \leq 3$ , одержимо

$$\begin{aligned} L &= 5x_1 + 4x_2 \rightarrow \max, \\ x_1 + x_2 &\leq 5, \\ 10x_1 + 6x_2 &\leq 45, \\ x_1 &\leq 3, \\ x_j &\geq 0, \quad x_j - \text{цілі}, \quad j = \overline{1,2}. \end{aligned}$$

Використовуючи функцію Excel «Пошук рішення», отримаємо оптимальний цілочисловий план

$$x_1=3; x_2=2; L=23.$$

Якість отриманого цілочислового розв'язку оцінити неможливо, тому що розв'язок другої задачі може призвести до більшого значення цільової функції. Це треба перевірити. Сформуємо другу задачу, додавши до неї обмеження  $x_1 \geq 4$ :

$$L = 5x_1 + 4x_2 \rightarrow \max,$$

$$x_1 + x_2 \leq 5,$$

$$10x_1 + 6x_2 \leq 45,$$

$$x_1 \geq 4,$$

$$x_j \geq 0, \quad x_j - \text{цілі}, \quad j = \overline{1,2}.$$

В оптимальному розв'язку другої задачі змінна  $x_2$  не є цілим числом:

$$x_1=4; x_2=0,83; L=23,33.$$

Оскільки значення змінної  $x_2=0,83$  не є цілим числом, другу задачу необхідно досліджувати далі, розділивши її у свою чергу ще на дві задачі й т.д. Проілюструємо хід розв'язання схемою, наведеною на рисунку 7.2.

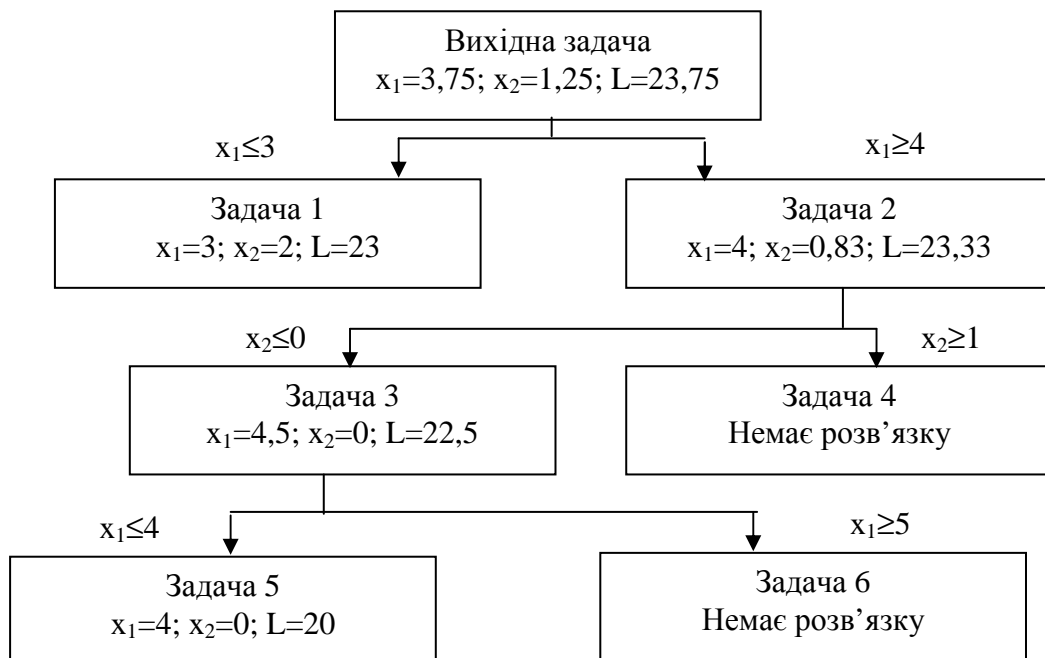


Рисунок 7.2 – Схема методу віток і границь

У ході розв'язання можна було спочатку як змінну розгалуження прийняти  $x_2$ . При цьому наступні обчислення можуть значно відрізнятись.

Очевидним недоліком алгоритму методу віток і границь при розв'язанні задач великої розмірності є необхідність перебрати занадто велику кількість варіантів, перш ніж буде знайдений оптимальний план.

### Контрольні запитання

1. Які основні проблеми виникають при розв'язанні дискретних задач?
2. Сформулюйте задачу про ранець.
3. Які економіко-математичні моделі можуть бути зведені до задачі про комівояжера?

4. Наведіть приклади моделей з розривними цільовими функціями.
5. Який принцип використовується для побудови правильного відсікання в методі Гоморі?
6. Яку роль відіграє алгоритм двоїстого симплекс-методу при розв'язанні цілочислової лінійної задачі методом Гоморі?
7. Перелічіть принципові ідеї, що лежать в основі методу віток і границь.
8. Як провадиться побудова відсікання при розв'язанні цілочислової лінійної задачі за методом віток і границь?
9. Опишіть схему розв'язання цілочислової задачі лінійного програмування за методом віток і границь.
10. За рахунок яких перетворень вдається побудувати сполучений базис при додаванні обмеження, що відтинає?

## ТЕМА 8

### ЗАДАЧІ НЕЛІНІЙНОГО ПРОГРАМУВАННЯ. ОСНОВНІ МЕТОДИ ЇХ РОЗВ'ЯЗАННЯ І АНАЛІЗУ

#### 8.1 Постановка задачі нелінійного програмування (ЗНП)

Залежності між керованими змінними далеко не завжди можна описати за допомогою адекватної лінійної моделі. Наприклад, у лінійних моделях ціна товару вважається незалежною від кількості реалізованого продукту, у той час як вона може залежати від обсягу партії товару. З приводу технологічних обмежень можна помітити, що витрата певних видів сировини і ресурсів відбувається не лінійно, а стрибкоподібно (залежно від обсягу виробництва). Спроби врахувати ці фактори приводять до формулювання більш загальних і складних оптимізаційних задач. Вивчення методів їх розв'язання складає предмет наукової області, що одержала назву **нелінійного програмування**.

Для задач нелінійного програмування не існує універсального методу розв'язання, тому щораз необхідно доводити існування розв'язку задачі, а також його єдиність. Відомі точні методи розв'язання нелінійних задач, але їхні алгоритми є трудомісткими навіть для сучасного програмного забезпечення ЕОМ. На практиці частіше користуються наближеними методами, проблема яких пов'язана з пошуком локальних і глобальних оптимумів. Більшість наближених методів дозволяють визначити локальний оптимум. Визначивши всі локальні оптимуми й порівнявши їх, можна знайти глобальний оптимум. Але такий підхід не є ефективним для практичних розрахунків. Слід зазначити, що якщо в задачах лінійного програмування оптимальний розв'язок завжди перебував на границі області обмежень, то у задачі нелінійного програмування він може перебувати також й усередині цієї області.

У класичній теорії оптимізації для пошуку точок максимуму й мінімуму (екстремальних точок) функцій, як при відсутності так і при наявності обмежень на змінні, використовується апарат диференціального обчислення. Екстремальна точка функції  $f(x)$  визначає або її максимальне, або мінімальне

значення. З математичної точки зору точка  $x_0=(x_1, x_2, \dots, x_n)$  є точкою максимуму, якщо значення функції  $f$  в оточенні точки  $x_0$  не перевищують  $f(x_0)$ . На рисунку 8.1 показані точки максимуму та мінімуму функції однієї змінної  $f(x)$ . Точки  $x_1, x_2, x_3, x_4$  і  $x_6$  складають множину екстремальних точок функції  $f(x)$ . Причому точка  $x_6$  є точкою **глобального (абсолютного)** максимуму, тому що  $f(x) = \max\{f(x_1), f(x_3), f(x_6)\}$ , а точки  $f(x_1)$  і  $f(x_3)$  – точками **локального (відносного)** максимуму. Необхідною умовою існування екстремуму є рівність нулю похідних від  $f(x)$ . Але похідні дорівнюють нулю також і в точках перегину функції  $f(x)$  у випадку однієї змінної (точка  $x_5$ ), і в сідлових точках у випадку функції двох змінних. Тому рівність нулю похідних від  $f(x)$  є необхідною, але недостатньою умовою існування екстремуму, а точки, у яких виконується дана умова, називаються **стаціонарними**.

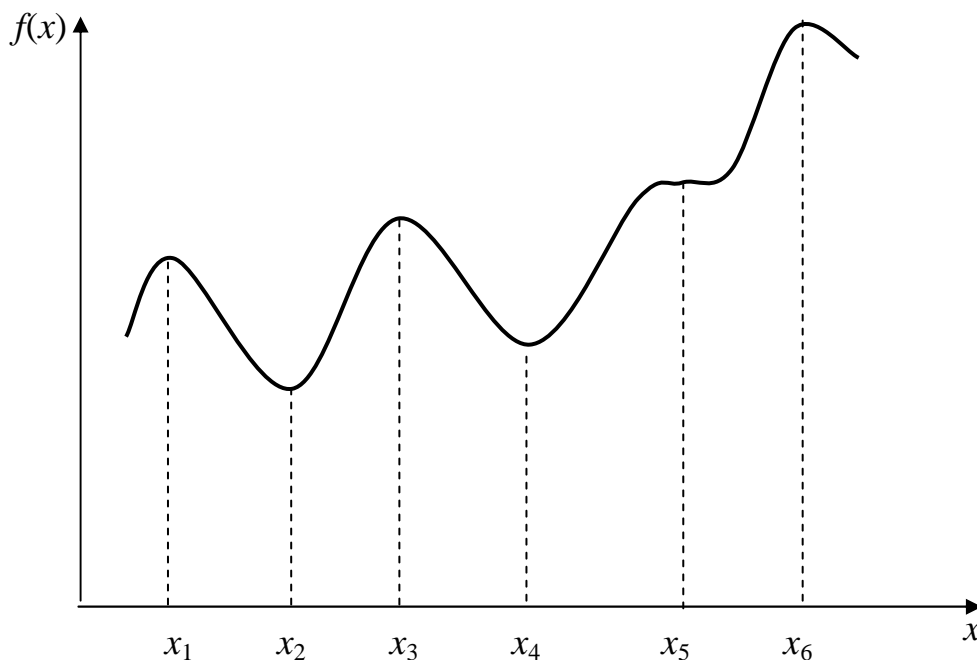


Рисунок 8.1 – Точки максимуму і мінімуму функції

Більшість методів, використовуваних для розв’язання задач нелінійного програмування, базуються на теорії диференціального обчислення. Серед них розрізняють прямі й непрямі методи. Прямими методами пошук оптимуму ведеться в напрямку найшвидшого зростання або убуття цільової функції. До таких методів належать градієнтні методи, зокрема, метод найшвидшого спуску і метод припустимих напрямків. Непрямі методи припускають перетворення вихідної задачі до виду, що дозволяє спростити пошук екстремуму цільової функції. До них належить квадратичне програмування й сепарабельне програмування, а також геометричне й стохастичне програмування.

Загальна задача нелінійного програмування (ЗНП) визначається як задача знаходження максимуму (або мінімуму) цільової функції  $f(x_1, x_2, \dots, x_n)$  на множині  $D$ , зумовленій системою обмежень

$$\begin{aligned} g_1(x_1, x_2, \dots, x_n) &\leq 0, \\ g_2(x_1, x_2, \dots, x_n) &\leq 0, \\ &\dots\dots\dots \end{aligned}$$

$$\begin{aligned} g_i(x_1, x_2, \dots, x_n) &\leq 0, \\ g_{i+1}(x_1, x_2, \dots, x_n) &\leq 0, \\ &..... \\ g_m(x_1, x_2, \dots, x_n) &= 0, \\ x_j &\geq 0, \quad j = \overline{1, n}, \end{aligned} \tag{8.1}$$

де хоча б одна з функцій  $f$  або  $g_i$  є нелінійною.

Очевидно, що питання про тип оптимізації не є принциповим. Тому, для визначеності, надалі будемо розглядати задачі максимізації.

Як і в ЗЛП, вектор  $x^*=(x_1, x_2,..., x_n) \in D$  називається припустимим планом, а якщо для будь-якого  $x \in D$  виконується нерівність  $f(x^*) \geq f(x)$ , то  $x^*$  називають оптимальним планом. У цьому випадку  $x^*$  є точкою глобального максимуму.

З погляду економічної інтерпретації  $f(x)$  може розглядатися як дохід, що одержує підприємство при плані випуску  $x$ , а  $g_i(x) < 0$  як технологічні обмеження на можливості випуску продукції.

Набір обмежень, які визначають множину  $D$ , при необхідності завжди можна звести або до системи, що складається з одних нерівностей, або, додавши фіктивні змінні, до системи рівнянь. Перелічимо властивості ЗНП, які істотно ускладнюють процес їхнього розв'язання в порівнянні із задачами лінійного програмування.

1. Множина припустимих планів  $D$  може мати дуже складну структуру (наприклад, бути неопуклою або незв'язною).

2. Глобальний максимум (мінімум) може досягатися як усередині множини  $D$ , так і на її границях (де він, загалом кажучи, буде не збігатися з жодним з локальних екстремумів).

3. Цільова функція  $f$  може бути недиференційованою, що утруднює застосування класичних методів математичного аналізу.

У чинність названих факторів задачі нелінійного програмування настільки різноманітні, що для них не існує загального методу розв'язання.

## 8.2 Класичний метод оптимізації з використанням множників Лагранжа

Одним з найбільш загальних підходів до розв'язання задач пошуку екстремуму функції при наявності сполучних обмежень на її змінні (задача умовної оптимізації) є метод Лагранжа. Багатьом студентам він має бути відомим з курсу диференціального обчислення. Ідея даного методу полягає у зведенні задачі пошуку умовного екстремуму цільової функції

$$f(x_1, x_2, \dots, x_n) \quad (8.2)$$

на множині припустимих значень  $D$ , описуваний системою рівнянь

$$\begin{aligned} g_1(x_1, x_2, \dots, x_n) &= 0, \\ g_2(x_1, x_2, \dots, x_n) &= 0, \\ &\dots\dots\dots \\ g_m(x_1, x_2, \dots, x_n) &= 0 \end{aligned} \quad (8.3)$$

до задачі безумовної оптимізації функції



$$\Phi(x, \lambda) = f(x_1, x_2, \dots, x_n) + \lambda_1 g_1(x_1, x_2, \dots, x_n) + \dots + \lambda_m g_m(x_1, x_2, \dots, x_n), \quad (8.4)$$

де  $\lambda_i$  - вектор додаткових змінних, називаних невизначеними множниками Лагранжа.

Рівняння (8.3) називаються рівняннями зв'язку, а функція  $\Phi(x, \lambda)$  - функцією Лагранжа.

Для функції  $\Phi(x, \lambda)$  вирішують систему рівнянь

$$\begin{cases} \frac{\partial \Phi(x, \lambda)}{\partial x_j} = 0, (j = \overline{1, n}), \\ \frac{\partial \Phi(x, \lambda)}{\partial \lambda_i} = g_i(x) = 0, (i = \overline{1, m}) \end{cases} \quad (8.5)$$

щодо змінних  $x$  і  $\lambda$ .

Метод Лагранжа складається з наступних етапів.

1. Складання функції Лагранжа  $\Phi(x, \lambda)$ .

2. Знаходження частинних похідних

$$\frac{\partial \Phi(x, \lambda)}{\partial x_j}, j = \overline{1, n} \quad \text{і} \quad \frac{\partial \Phi(x, \lambda)}{\partial \lambda_i}, i = \overline{1, m}.$$

3. Розв'язання системи рівнянь (8.5) щодо змінних  $x$  і  $\lambda$ .

4. Дослідження точок, що задовольняють системі (8.5), на максимум (мінімум) за допомогою достатньої ознаки екстремуму.

Наявність останнього (четвертого) етапу пояснюється тим, що розглянутий алгоритм виконує необхідну, але не достатню умову екстремуму. Для визначення достатніх ознак умовного екстремуму і його типу існує спеціальний алгоритм, як правило, важко застосовний на практиці.

Основне практичне значення методу Лагранжа полягає в тому, що він дозволяє перейти від умовної оптимізації до безумовної. Проте задача розв'язання системи рівнянь (8.5), до якої збігається даний метод, у загальному випадку не простіше вихідної проблеми пошуку екстремуму (8.2)-(8.3). Методи, що припускають таке розв'язання, називаються непрямими. Вони можуть бути застосовані для досить вузького класу задач, для яких вдається одержати лінійну або таку, що зводиться до лінійної систему рівнянь (8.5). Їхнє застосування пояснюється необхідністю одержати розв'язок екстремальної задачі в аналітичній формі. При розв'язанні конкретних практичних задач звичайно використовують прямі методи, засновані на ітеративних процесах обчислення й порівнянні значень оптимізуємих функцій.

### 8.3 Опукле програмування

Основний недолік методів нелінійного програмування полягає в тому, що за їхньою допомогою не вдається знайти глобальний екстремум за наявності кількох локальних екстремумів. Теоретично нелінійне програмування розроблене тільки для одного окремого випадку опуклих функцій, і відповідно цей розділ названий опуклим програмуванням.

Функція  $f(x_1, x_2, \dots, x_n)$  називається опуклою в області  $D$ , якщо для будь-яких двох точок  $x^{(1)}, x^{(2)} \in D$  і будь-якого  $\lambda \in [0, 1]$  виконується нерівність

$$f((1-\lambda)x^{(1)} + \lambda x^{(2)}) \leq (1-\lambda)f(x^{(1)}) + \lambda f(x^{(2)}), \quad (8.6)$$

якщо ж

$$f((1-\lambda)x^{(1)} + \lambda x^{(2)}) \geq (1-\lambda)f(x^{(1)}) + \lambda f(x^{(2)}), \quad (8.7)$$

то функція називається увігнутою.

Геометричний зміст понять опуклості та увігнутості для функції однієї змінної поданий на рисунку 8.2. З нього видно, що графік опуклої функції лежить нижче відрізка, що з'єднує точки  $(x^{(1)}, f(x^{(1)}))$  і  $(x^{(2)}, f(x^{(2)}))$ , а графік увігнутої – вище.

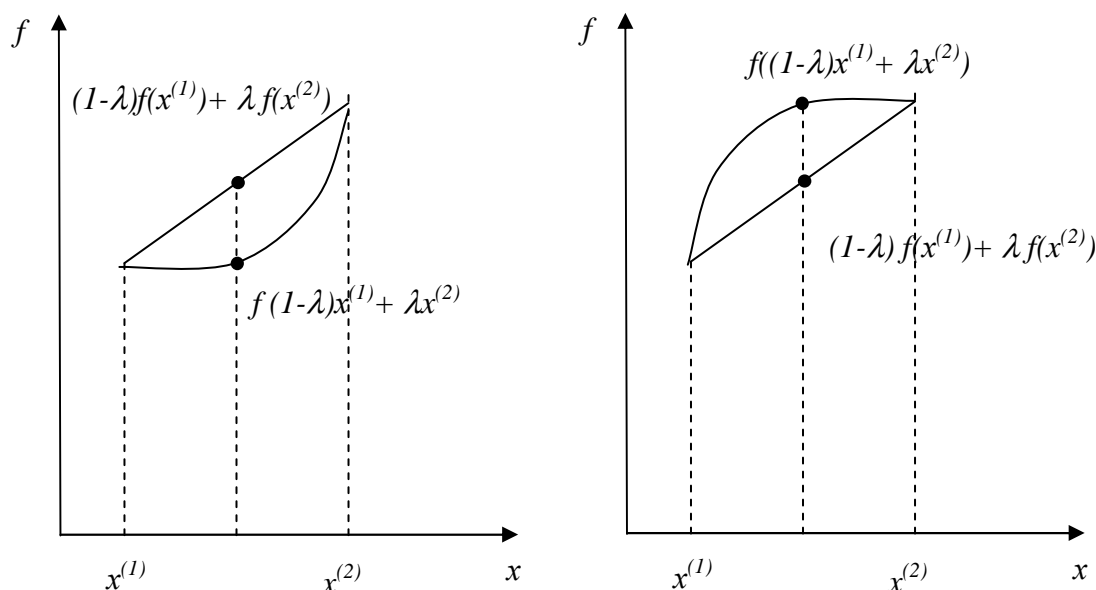


Рисунок 8.2 - Графіки опуклої та увігнутої функцій

Можна довести, що достатньою умовою опуклості функції  $f(x_1, x_2, \dots, x_n)$  є позитивна визначеність матриці

$$H(x) = \left( \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{n \times n}, \quad (8.8)$$

називаної матрицею Гессе, у всіх точках  $x \in D$ . Відповідно, достатньою умовою увігнутості є негативна визначеність матриці Гессе. Зокрема, для функцій однієї змінної достатньою умовою опуклості (увігнутості) є виконання нерівності  $f''(x) \geq 0$  ( $f''(x) \leq 0$ ).

Як видно з геометричної інтерпретації, для опуклої функції локальний екстремум, якщо він існує, збігається із глобальним. Для функцій багатьох змінних точка  $x_0$  являє собою вектор  $f'(x)$  – вектор перших похідних (градієнт) функції  $f(x)$ , а матриця Гессе (гессіан)  $f''(x)$  – симетричну матрицю других частинних похідних функції  $f(x)$ . Характер стаціонарної точки  $x_0$  зв'язаний зі знаковизначеністю матриці Гессе  $f''(x^*)$ .

Знаковизначеність матриці  $A$  залежить від знаків квадратичної форми

$$Q(x) = X^T A X = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j. \quad (8.9)$$

Квадратична форма  $Q(x)$  є позитивно визначеною (напіввизначеною), якщо значення всіх кутових мінорів визначника  $|A|$  додатні (невід'ємні). У цьому випадку матриця  $A$  називається позитивно визначеною (напіввизначеною).  $k$ -м кутовим мінором визначника матриці  $A_{n \times n}$  називається визначник виду

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \dots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{vmatrix}, \quad k = \overline{1, n}. \quad (8.10)$$

Квадратична форма  $Q(x)$  є негативно визначеною, якщо значення  $k$ -х кутових мінорів визначника  $|A|$  відмінні від нуля та мають знак  $(-1)^k$ . У цьому випадку матриця  $A$  називається негативно визначеною.

Квадратична форма  $Q(x)$  є негативно напіввизначеною, якщо значення  $k$ -х кутових мінорів визначника  $|A|$  дорівнюють нулю або мають знак  $(-1)^k$ .

Розглянемо приклад. Нехай є функція

$$f(x_1, x_2, x_3) = x_1 + 2x_3 + x_2x_3 - x_1^2 - x_2^2 - x_3^2,$$

необхідно визначити її екстремум. Визначимо градієнт функції  $f(x)$  (необхідна умова екстремуму)

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 1 - 2x_1 = 0, \\ \frac{\partial f}{\partial x_2} &= x_1 - 2x_2 = 0, \\ \frac{\partial f}{\partial x_3} &= 2 + x_2 - 2x_3 = 0, \end{aligned}$$

звідки одержимо стаціонарну точку  $x_0 = \left(\frac{1}{2}, \frac{2}{3}, \frac{4}{3}\right)$ . Визначимо характер стаціонарної точки. Для цього перевіримо достатність умови, склавши матрицю Гессе

$$H(x_0) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} \end{pmatrix}_{x_0} = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix}$$

і кутові мінори матриці  $H(x_0)$

$$M_1 = -2; M_2 = (-2) \cdot (-2) - 0 \cdot 0 = 4; M_3 = -2 \cdot [(-2) \cdot (-2) - 1 \cdot 1] = -6.$$

Таким чином, матриця  $H(x_0)$  є негативно визначеною, звідки випливає, що функція  $f(x)$  – увігнута, а отже стаціонарна точка  $x_0 = \left(\frac{1}{2}, \frac{2}{3}, \frac{4}{3}\right)$  є точкою максимуму.

## 8.4 Необхідні й достатні умови існування сідлової точки.

### Теорема Куна-Таккера

Зупинимось на деяких фундаментальних моментах теорії нелінійного програмування. Відправною точкою для них є поширення методу Лагранжа на розв'язання ЗНП із обмеженнями у формі нерівностей:

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &\rightarrow \max, \\ g_1(x_1, x_2, \dots, x_n) &\leq 0, \\ g_2(x_1, x_2, \dots, x_n) &\leq 0, \\ &\dots\dots\dots \\ g_m(x_1, x_2, \dots, x_n) &\leq 0 \end{aligned} \quad (8.11)$$

Визначимо функцію Лагранжа:

$$\Phi(x, \lambda) = f(x_1, x_2, \dots, x_n) - \sum_{i=1}^m \lambda_i g_i(x_1, x_2, \dots, x_n). \quad (8.12)$$

**Пара векторів  $(x, \lambda)$  називається сідловою точкою функції  $\Phi(x, \lambda)$  у певній області, якщо для будь-яких  $x$  і  $\lambda$ , що належать цієї області,**

$$\Phi(x, \lambda^*) \leq \Phi(x^*, \lambda^*) \leq \Phi(x^*, \lambda). \quad (8.13)$$

Нерівності (8.13) називають нерівностями сідлової точки.

Як приклад розглянемо функцію  $\Phi(x, \lambda) = -x^2 + \lambda^2$ . Сідловою точкою для цієї функції є точка  $(0, 0)$ . Справді,  $\Phi(0, 0) = 0$ ,  $\Phi(x, 0) = -x^2$ ,  $\Phi(0, \lambda) = \lambda^2$ , а для будь-яких  $x$  і  $\lambda$  виконуються нерівності  $-x^2 \leq 0$  і  $0 \leq \lambda^2$ .

На рисунку 8.3 зображений графік функції  $\Phi(x, \lambda)$  (гіперболічний параболоїд), і, як видно, в оточенні точки  $(0, 0)$  він дійсно за формою нагадує сідло, чим і пояснюється походження відповідного терміна.

Центральне місце в теорії нелінійного програмування займає **теорема Куна-Таккера**, що зв'язує розв'язання ЗНП із наявністю сідлової точки у відповідної функції Лагранжа.

**Теорема 8.1.** (Достатня умова екстремуму).

Якщо  $(x^*, \lambda^*)$  - сідлова точка функції Лагранжа в області  $x \in X \supseteq D$ ,  $\lambda \geq 0$ , то  $x^*$  є оптимальним планом задачі (8.11), причому справедливе так зване правило нетвердості, що доповнює (умова Слейтера):

$$\sum_{i=1}^m \lambda_i^* g_i(x^*) = 0. \quad (8.14)$$

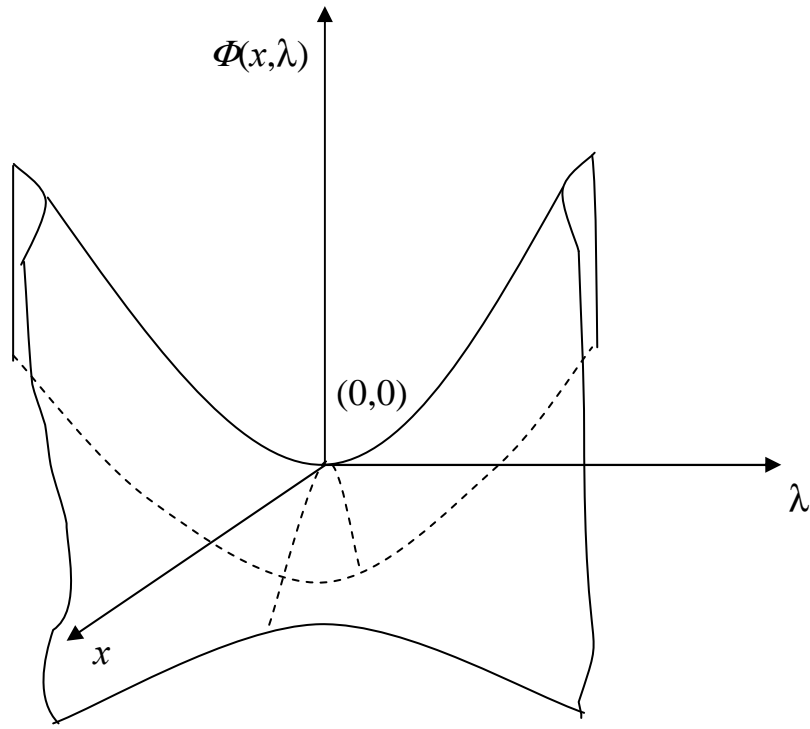


Рисунок 8.3 - Графік функції  $\Phi(x, \lambda)$

Доказ.

Скористаємося визначенням сідлової точки й запишемо

$$f(x) - \sum_{i=1}^m \lambda_i^* g_i(x) \leq f(x^*) - \sum_{i=1}^m \lambda_i^* g_i(x^*) \leq f(x^*) - \sum_{i=1}^m \lambda_i g_i(x^*) \quad (8.15)$$

при всіх  $x \in X$ ,  $\lambda \geq 0$ . Із другої нерівності в (8.15) випливає, що

$$\sum_{i=1}^m \lambda_i g_i(x^*) \leq \sum_{i=1}^m \lambda_i^* g_i(x^*) \text{ для } \lambda \geq 0. \quad (8.16)$$

Проте (8.16) може мати місце тільки тоді, коли  $g(x^*) \leq 0$  при всіх  $i = \overline{1, m}$ . Дійсно, якщо існує таке  $k$ , що  $g_k(x^*) > 0$ , то поклавши  $\lambda_i = 0$  для всіх  $i \neq k$  і вибравши досить велике  $\lambda_k > 0$ , можна домогтися того, що значення  $\sum_{i=1}^m \lambda_i g_i(x^*) = \lambda_k g_k(x^*)$  виявиться більшим за постійний вираз  $\sum_{i=1}^m \lambda_i^* g_i(x^*)$ .

З того, що для всіх  $i = \overline{1, m}$  виконуються нерівності  $g_i(x^*) \leq 0$ , випливає, що  $x^*$  є припустимим планом задачі (8.11).

Якщо до лівої частини нерівності (8.16) підставити значення  $\lambda_i = 0$  для всіх  $i = \overline{1, m}$ , то одержимо

$$\sum_{i=1}^m \lambda_i^* g_i(x^*) \geq 0.$$

Разом з тим з того що,  $g_i(x^*) \leq 0$  і  $\lambda_i^* \geq 0$  випливає оцінка

$$\sum_{i=1}^m \lambda_i^* g_i(x^*) \leq 0.$$

Спільний розгляд останніх двох нерівностей призводить до правила нетвердості, що доповнює, у точці  $x^*$ :

$$\sum_{i=1}^m \lambda_i^* g_i(x^*) = 0.$$

Тоді на підставі лівої частини нерівності сідлової точки (8.15) маємо, що для всіх  $x \in X$  (у тому числі й для  $x \in D$ )

$$f(x^*) \geq f(x) - \sum_{i=1}^m \lambda_i^* g_i(x).$$

За умовою ЗНП для будь-яких  $x \in D$  вірними є нерівності  $g_i(x^*) \leq 0$ , що в сполученні з умовою  $\lambda_i^* = 0$  дозволяє записати

$$\sum_{i=1}^m \lambda_i^* g_i(x) \leq 0.$$

Виходить,

$$f(x) - \sum_{i=1}^m \lambda_i^* g_i(x) \geq f(x^*).$$

Остаточно одержуємо, що для будь-яких  $x \in D$  справедливим є співвідношення  $f(x^*) \geq f(x)$ , тобто  $x^*$  - оптимальний план задачі (8.11).

Значення теореми Куна-Таккера полягає в тому, що вона дозволяє зв'язати процес розв'язання оптимізаційної задачі з пошуком сідлових точок функції Лагранжа, тобто з максимізацією цієї функції за  $x$  і мінімізацією за  $\lambda$ .

Нехай  $F(x)$  є функцією, що ставить у відповідність кожному значенню  $x$  мінімальне значення функції  $\Phi(x, \lambda)$  за  $\lambda$ :

$$F(x) = \min_{\lambda \geq 0} \Phi(x, \lambda)$$

і за аналогією

$$G(\lambda) = \max_{x \in X} \Phi(x, \lambda).$$

Розглянемо задачу відшукування максимуму функції  $F(x)$

$$F(x) = \min_{\lambda \geq 0} \Phi(x, \lambda) \rightarrow \max, \quad x \in X \quad (8.17)$$

і задачу мінімізації  $G(\lambda)$

$$G(\lambda) = \max_{x \in X} \Phi(x, \lambda) \rightarrow \min, \quad \lambda \geq 0. \quad (8.18)$$

Очевидно, що

$$F(x) = \min_{\lambda \geq 0} \Phi(x, \lambda) = \begin{cases} f(x), & x \in D, \\ \infty, & x \notin D \end{cases}.$$

Звідси випливає, що максимум  $F(x)$  знаходиться в припустимій області  $D$  і збігається з максимумом цільової функції  $f(x)$  задачі (8.11):

$$\max_{x \in X} F(x) = \max_{x \in X} \min_{\lambda \geq 0} \Phi(x, \lambda) = \max_{x \in D} f(x).$$

Таким чином, задача (8.17), у певному сенсі, рівносильна (8.11). Аналогічні висновки можуть бути отримані і для (8.18). Задачі (8.17) і (8.18) утворюють двоїсту пару. Дане відношення є узагальненням відносини

подвійності для задач лінійного програмування. Відповідно, за певних умов пара двоїстих задач нелінійного програмування має властивості, аналогічні властивостям двоїстих лінійних задач. Зокрема, при будь-яких  $x \in X, \lambda \geq 0$

$$F(x) \leq G(\lambda). \quad (8.19)$$

Умова (8.19) знаходить широке застосування при побудові оцінок в ітеративних методах розв'язання оптимізаційних задач. Наприклад, якщо є можливість приблизно вирішити пряму і двоїсту задачі і одержати послідовності наближень, то за допомогою нерівностей виду

$$f(x^{(k)}) \leq f(x^*) \leq G(\lambda^{(k)})$$

можна визначити момент зупинки обчислювальної процедури.

На закінчення відзначимо, що можливим є варіант виведення виразів для цільових функцій і обмежень пари двоїстих задач лінійного програмування із загального визначення відносини подвійності для нелінійних задач. Також відзначимо, що в процесі формування нелінійних двоїстих задач існує велика неоднозначність: їхній вид можна варіювати, включаючи до множини  $X$  частину обмежень  $g_i(x) \leq 0$ .

## 8.5 Деякі методи розв'язання задач НЛП

**Гradientні методи розв'язання задач безумовної оптимізації.** Провідне місце серед прямих методів розв'язання екстремальних задач займає gradientний метод (точніше, сімейство gradientних методів) пошуку стаціонарних точок диференційованої функції. Нагадаємо, що стаціонарною називається точка, у якій  $\nabla f(x) = 0$  і яка відповідно до необхідної умови оптимальності є «підозрілою» на наявність локального екстремуму. Таким чином, застосовуючи gradientний метод, знаходять множину точок локальних максимумів (або мінімумів), серед яких визначають максимум (або мінімум) глобальний.

Ідея даного методу заснована на тому, що gradient функції вказує напрямок її найбільш швидкого зростання в оточенні тієї точки, в якій він обчислений. Тому, якщо з певної поточної точки  $x^{(1)}$  переміщатися в напрямку вектора  $\nabla f(x^{(1)})$ , функція  $f$  зростатиме, принаймні, у певному оточенні  $x^{(1)}$ . Отже, для точки  $x^{(2)} = x^{(1)} + k \nabla f(x^{(1)})$ , ( $k > 0$ ), що лежить у такому оточенні, справедлива нерівність  $f(x^{(1)}) \leq f(x^{(2)})$ . Продовжуючи цей процес, ми поступово будемо наближатися до точки деякого локального максимуму.

Однак як тільки визначається напрямок руху, відразу ж встає питання, як далеко треба рухатися в цьому напрямку, тобто виникає проблема вибору кроку  $r$  у рекурентній формулі

$$x^{(k+1)} = x^{(k)} + r \nabla f(x^{(k)}), \quad (8.20)$$

що задає послідовність точок, які прагнуть до точки максимуму.

Залежно від способу її розв'язання розрізняють різні варіанти gradientного методу. Зупинимось на найбільш відомих з них.

**Метод найшвидшого спуска.** Назву методу можна було б розуміти буквально, якби мова йшла про мінімізацію цільової функції. Проте, за традицією така назва використовується і при розв'язанні задачі на максимум.

Нехай  $f(x) = f(x_1, x_2, \dots, x_n)$  - диференційована функція, а  $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})$  - деяка поточна точка. Будь-яких загальних рекомендацій щодо вибору вихідної точки (початкового наближення)  $x^{(0)}$  не існує, але за можливістю вона має знаходитись близько від шуканого оптимального плану  $x^*$ . Якщо  $x^{(k)}$  - нестационарна точка, то при русі в напрямку  $\nabla f(x^{(k)})$  функція  $f(x)$  на певному проміжку обов'язково зростатиме. Звідси виникає необхідність такого вибору кроку, щоб рух у зазначеному напрямку тривав доти, поки зростання не припиниться. Виразимо залежність значення  $f(x)$  від крокового множника  $r > 0$ , покладаючи  $x = x^{(k)} + r \nabla f(x^{(k)})$

$$f(x) = f(x^{(k)} + r \nabla f(x^{(k)})) = \varphi(r), \quad (8.21)$$

або в координатній формі,

$$\varphi(r) = f\left(x_1^{(k)} + r \frac{\partial f(x^{(k)})}{\partial x_1}, \dots, x_n^{(k)} + r \frac{\partial f(x^{(k)})}{\partial x_n}\right). \quad (8.22)$$

Щоб домогтися найбільшого з можливих значень  $f$  при русі у напрямку  $\nabla f(x^{(k)})$ , потрібно вибрати таке значення  $r$ , що максимізує функцію  $\varphi(r)$ . Для обчислення  $r$  використовується необхідна умова екстремуму  $\frac{d\varphi(r)}{dr} = 0$ .

Помітимо, що якщо для будь-якого  $r > 0$   $\frac{d\varphi(r)}{dr} > 0$ , то функція  $f(x)$  не обмежена зверху (тобто не має максимуму). У протилежному випадку, на підставі (8.22) одержуємо

$$\frac{d\varphi(r)}{dr} = \frac{\partial f(x)}{\partial x_1} \times \frac{dx_1}{dr} + \dots + \frac{\partial f(x)}{\partial x_n} \times \frac{dx_n}{dr}, \quad (8.23)$$

що, у свою чергу, дає

$$\frac{d\varphi(r)}{dr} = \frac{\partial f(x)}{\partial x_1} \times \frac{\partial f(x^{(k)})}{\partial x_1} + \dots + \frac{\partial f(x)}{\partial x_n} \times \frac{\partial f(x^{(k)})}{\partial x_n} = \nabla f(x) \nabla f(x^{(k)}). \quad (8.24)$$

Якщо вважати, що наступна точка  $x^{(k+1)}$  відповідає оптимальному значенню, то в ній має виконуватися умова  $\frac{d\varphi(r)}{dr} = 0$ , і  $r$  треба знаходити з умови  $\nabla f(x^{(k+1)}) \nabla f(x^{(k)}) = 0$  або

$$\nabla f(x^{(k)} + r \nabla f(x^{(k)})) \nabla f(x^{(k)}) = 0. \quad (8.25)$$

Умова (8.25) означає рівність нулю скалярного добутку градієнтів функції  $f$  в точках  $x^{(k+1)}$  і  $x^{(k)}$ . Геометрично вона може бути інтерпретованою як перпендикулярність векторів градієнтів функції  $f$  у зазначених точках. Продовжуючи геометричну інтерпретацію методу найшвидшого спуска, відзначимо, що в точці  $x^{(k+1)}$  вектор  $\nabla f(x^{(k+1)})$ , будучи градієнтом, перпендикулярний до лінії рівня, що проходить через дану точку. Відповідно вектор  $\nabla f(x^{(k)})$  є дотичним до цієї лінії. Таким чином, рух у напрямку



градієнта  $\nabla f(x^{(k)})$  треба продовжувати доти, поки він перетинає лінії рівня оптимізуємої функції.

Після того як точку  $x^{(k+1)}$  знайдено, вона стає поточною для чергової ітерації. На практиці ознакою досягнення стаціонарної точки служить досить мала зміна координат точок, розглянутих на послідовних ітераціях. Одночасно із цим координати вектора  $\nabla f(x^{(k)})$  повинні бути близькі до нуля.

**Метод дроблення кроку.** Для знаходження кроку  $r$  у методі найшвидшого спуску потрібно вирішити рівняння (8.25), що може виявитися досить складним. Тому часто обмежуються «підбором» такого значення  $r$ , що  $\varphi(r) > \varphi(0)$ . Для цього задаються деяким початковим значенням  $r_1$  (наприклад,  $r_1=1$ ) і перевіряють умову  $\varphi(r_1) > \varphi(0)$ . Якщо вона не виконується, то покладають

$$r_2 = \frac{r_1}{2}$$

і т. д. доти, поки не вдається знайти підходящий крок, з яким переходять до наступної точки  $x^{(k+1)}$ . Критерій завершення алгоритму буде таким самим, як і в методі найшвидшого спуску.

**Квадратичне програмування (КП).** До задач квадратичного програмування належить спеціальний клас задач нелінійного програмування, для яких цільова функція  $f(x)$  квадратична, а всі обмеження - лінійні. Застосувавши до цієї задачі теорему Куна-Таккера, одержують умову для оптимального розв'язку у вигляді системи лінійних рівнянь, вирішити які можна симплекс-методом. У матричному виді задача квадратичного програмування формулюється в такий спосіб.

Максимізувати (мінімізувати) цільову функцію  $f = \overline{C}\overline{X} + \overline{X}^T \overline{D}\overline{X}$  при обмеженнях

$$\overline{A}\overline{X} \leq b, \quad \overline{X} \geq 0, \quad (8.26)$$

де

$$\overline{X} = (x_1, x_2, \dots, x_n)^T,$$

$$\overline{C} = (c_1, c_2, \dots, c_n),$$

$$b = (b_1, b_2, \dots, b_m)^T,$$

$$\overline{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix},$$

$$\overline{D} = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \dots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{pmatrix}.$$

Функція  $\bar{X}^T \bar{D} \bar{X}$ , де  $D$  – симетрична матриця, що є квадратичною формою (8.9). Матриця  $D$  є негативно визначеною в задачі максимізації й позитивно визначеною – у задачі мінімізації. Це означає, що функція  $f$  є строго опуклою за змінними  $\bar{X}$  у задачі мінімізації й строго увігнутою – у задачі максимізації. Обмеження задачі є лінійними, що гарантує опуклість області припустимих розв’язків.

Складемо функцію Лагранжа для задачі (8.26)

$$\Phi(\bar{X}, \bar{\lambda}) = \bar{C} \bar{X} + \bar{X}^T \bar{D} \bar{X} + \bar{\lambda}(b - \bar{A} \bar{X}). \quad (8.27)$$

Відповідно до теореми Куна-Таккера для існування сідлової точки необхідно й достатньо, щоб похідні функції Лагранжа за  $x_j$  були менше нуля, а за  $\lambda_i$  – більше нуля, маємо:

$$\frac{\partial \Phi(\bar{X}, \bar{\lambda})}{\partial x_j} = \bar{C} + 2\bar{D} \bar{X} - \bar{\lambda} \bar{A} \leq 0, \quad (8.28)$$

причому, якщо  $\frac{\partial \Phi}{\partial x_j} < 0$ , відповідні  $x_j$  дорівнюють нулю;

$$\frac{\partial \Phi(\bar{X}, \bar{\lambda})}{\partial \lambda_i} = b - \bar{A} \bar{X} \geq 0, \quad (8.29)$$

якщо  $\frac{\partial \Phi}{\partial \lambda_i} > 0$ , відповідні  $\lambda_i$  дорівнюють нулю.

Щоб нерівності (8.28) і (8.29) привести до рівностей, уведемо два допоміжних вектори:  $\bar{V} = (v_1, v_2, \dots, v_n) \geq 0$  і  $\bar{S} = (s_1, s_2, \dots, s_m) \geq 0$ , причому виберемо  $v_j > 0$ , якщо  $\frac{\partial \Phi}{\partial x_j} < 0$  й  $v_j = 0$ , якщо  $\frac{\partial \Phi}{\partial x_j} = 0$ , а також  $\lambda_i > 0$ , якщо  $\frac{\partial \Phi}{\partial \lambda_i} > 0$  й  $\lambda_i = 0$ , якщо  $\frac{\partial \Phi}{\partial \lambda_i} = 0$ , одержимо

$$\bar{C} + 2\bar{D} \bar{X} - \bar{\lambda} \bar{A} + \bar{V} = 0, \quad (8.30)$$

$$b - \bar{A} \bar{X} - \bar{S} = 0. \quad (8.31)$$

Порівнявши компоненти векторів  $\bar{X}$  і  $\bar{V}$ , а також  $\bar{\lambda}$  і  $\bar{S}$ , одержимо дві умови нетвердості, що доповнюють

$$\bar{X}^T \bar{V} = 0 \text{ і } \bar{\lambda}^T \bar{S} = 0. \quad (8.32)$$

З умов (8.32) випливає, що хоча б  $n$  змінних з  $\bar{X}$  і  $\bar{V}$ , а також  $m$  змінних з  $\bar{\lambda}$  і  $\bar{S}$  звертаються на нуль. Якщо існує оптимальний розв’язок задачі (8.26), то він є одним з базисних розв’язків системи (8.30), (8.31). Для знаходження припустимого базисного розв’язку використовується симплекс-метод.

Розглянемо приклад. Знайти

$$\max f(x_1, x_2) = \max(10x_1 + 20x_2 + x_1x_2 - 2x_1^2 - 2x_2^2)$$

при обмеженнях:

$$\begin{aligned} 8 - x_2 &\geq 0, \\ 10 - x_1 - x_2 &\geq 0, \\ x_1 &\geq 0, x_2 &\geq 0. \end{aligned}$$

Можна показати, що функція  $f(x_1, x_2)$  є увігнутою функцією. Складемо функцію Лагранжа

$$\Phi(x_1, x_2, \lambda_1, \lambda_2) = 10x_1 + 20x_2 + x_1x_2 - 2x_1^2 - 2x_2^2 + \lambda_1(8 - x_2) + \lambda_2(10 - x_1 - x_2).$$

Застосувавши теорему Куна-Таккера, одержимо умови існування сідлової точки:

$$\begin{aligned}\frac{\partial \Phi}{\partial x_1} &= 10 + x_2 - 4x_1 + \lambda_2 \leq 0, \\ \frac{\partial \Phi}{\partial x_2} &= 20 + x_1 - 4x_2 - \lambda_1 - \lambda_2 \leq 0, \\ \frac{\partial \Phi}{\partial \lambda_1} &= 8 - x_2 \geq 0, \\ \frac{\partial \Phi}{\partial \lambda_2} &= 10 - x_1 - x_2 \geq 0.\end{aligned}$$

Одержали систему лінійних нерівностей:

$$\begin{cases} 10 + x_2 - 4x_1 + \lambda_2 \leq 0, \\ 20 + x_1 - 4x_2 - \lambda_1 - \lambda_2 \leq 0, \\ 8 - x_2 \geq 0, \\ 10 - x_1 - x_2 \geq 0. \end{cases}$$

Уведемо вільні змінні  $\bar{V}$  й  $\bar{S}$ , що обертають систему нерівностей на систему лінійних рівнянь, вирішивши яку симплекс-методом, знайдемо оптимум

$$\begin{cases} 4x_1 - x_2 - \lambda_2 - v_1 = 10, \\ -x_1 + 4x_2 + \lambda_1 + \lambda_2 - v_2 = 20, \\ x_2 + s_1 = 8, \\ x_1 + x_2 + s_2 = 10. \end{cases}$$

Для знаходження припустимого базисного плану отриманої задачі скористаємося методом мінімізації нев'язань. Для цього побудуємо допоміжну задачу, увівши до обмежень 1 і 2 фіктивні змінні  $x_3$  і  $x_4$ , які входять до цільової функції допоміжної задачі з коефіцієнтами  $M$  (а інші змінні – з нульовими коефіцієнтами). Таким чином, допоміжна задача має вигляд

$$\begin{aligned}f^*(x) &= Mx_3 + Mx_4, \\ \begin{cases} 4x_1 - x_2 - \lambda_2 - v_1 + x_3 = 10, \\ -x_1 + 4x_2 + \lambda_1 + \lambda_2 - v_2 + x_4 = 20, \\ x_2 + s_1 = 8, \\ x_1 + x_2 + s_2 = 10. \end{cases}\end{aligned}$$

Складемо симплекс-таблицю 8.1.

Таблиця 8.1 – Вихідна симплекс-таблиця

Базис	$C_{баз}$	$C_j$	0	0	0	0	0	0	M	M	0	0
		$B$	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_7$	$A_8$	$A_9$	$A_{10}$
$A_7$	$M$	10	4	-1	0	-1	-1	0	1	0	0	0
$A_8$	$M$	20	-1	4	1	1	0	-1	0	1	0	0
$A_9$	0	8	0	1	0	0	0	0	0	0	1	0
$A_{10}$	0	10	1	1	0	0	0	0	0	0	0	1
$L_j$	$30M$		$3M$	$3M$	$M$	0	$-M$	$-M$	$M$	$M$	0	0
$\Delta_j$			$3M$	$3M$	$M$	0	$-M$	$-M$	0	0	0	0

Оскільки в першу чергу необхідно мінімізувати нев'язання, виведемо з базису вектор  $\bar{A}_7$  і введемо вектор  $\bar{A}_1$  (тим самим змінна  $x_3$  стане рівною нулю, а змінна  $x_1$  увійде до базису). Складемо нову симплекс-таблицю 8.2.

Таблиця 8.2 – Мінімізація нев'язань

Базис	$C_{баз}$	$C_j$	0	0	0	0	0	0	$M$	$M$	0	0
		$B$	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_7$	$A_8$	$A_9$	$A_{10}$
$A_1$	0	10/4	1	-1/4	0	-1/4	-1/4	0	1/4	0	0	0
$A_8$	$M$	90/4	0	15/4	1	3/4	-1/4	-1	1/4	1	0	0
$A_9$	0	8	0	1	0	0	0	0	0	0	1	0
$A_{10}$	0	30/4	0	5/4	0	1/4	1/4	0	-1/4	0	0	1
$L_j$		90M/4	0	15M/4	M	3M/4	-M/4	-M	M/4	M	0	0
$\Delta_j$			0	15M/4	M	3M/4	-M/4	-M	3M/4	0	0	0

Тепер виведемо з базису вектор  $\bar{A}_8$  і введемо  $\bar{A}_2$

Таблиця 8.3 – Наступна симплекс-таблиця

Базис	$C_{баз}$	$C_j$	0	0	0	0	0	0	$M$	$M$	0	0
		$B$	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_7$	$A_8$	$A_9$	$A_{10}$
$A_1$	0	4	1	0	1/15	-1/5	-4/15	-1/15	4/15	1/15	0	0
$A_2$	0	6	0	1	4/15	3/15	-1/15	-4/15	1/15	4/15	0	0
$A_9$	0	2	0	0	-4/15	-1/5	1/15	4/15	-1/15	-4/15	1	0
$A_{10}$	0	0	0	0	-1/3	0	1/3	1/3	-1/3	-1/3	0	1
$L_j$		0	0	0	0	0	0	0	0	0	0	0
$\Delta_j$		0	0	0	0	0	0	0	-M	-M	0	0

Таким чином, отриманий припустимий базисний план вихідної задачі:

$$x_1=4; x_2=6; \lambda_1=0; \lambda_2=0; v_1=0; v_2=0; s_1=2; s_2=0.$$

Перевіримо виконання умов нетвердості, що доповнюють (7.32)  $\bar{X}^T \bar{V} = 0$  і  $\bar{\lambda}^T \bar{S} = 0$ :

$$x_1^* v_1 = 0; x_2^* v_2 = 0; \lambda_1^* s_1 = 0; \lambda_2^* s_2 = 0.$$

Оскільки умови нетвердості, що доповнюють, виконуються, отриманий план є оптимальним. Оптимальне значення цільової функції

$$f(x_1, x_2) = 10 \cdot 4 + 20 \cdot 6 + 4 \cdot 6 - 2 \cdot 4^2 - 2 \cdot 6^2 = 80.$$

### Контрольні запитання

1. За яких умов оптимізаційна задача може бути віднесена до класу нелінійних?
2. Наведіть приклад економічної моделі, що зводиться до задачі нелінійного програмування.
3. Перелічіть основні труднощі, що виникають у процесі розв'язання задачі нелінійного програмування.
4. Який зміст вкладається в поняття «умовна оптимізація»?
5. Для чого призначений метод множників Лагранжа й у чому він полягає?
6. Яка точка називається стаціонарною?
7. Які принципові етапи входять у градієнтні методи?

8. Для розв'язання яких задач призначений метод найшвидшого спуска й метод дроблення кроку?
9. Дайте визначення опуклої (увігнутої) функції.
10. Сформулюйте достатню умову опуклості (увігнутості) функції.
11. У чому полягає специфіка задач опуклого програмування?
12. Дайте визначення сідлової точки. Наведіть приклад функції, що має сідлову точку.
13. Сформулюйте необхідну й достатню умови теореми Куна-Таккера. Яке значення вони мають для розв'язання задач нелінійного програмування?
14. У чому полягає умова регулярності Слейтера? Поясніть її зміст.
15. Приведіть приклад пари двоїстих задач нелінійного програмування.
16. Які властивості пари нелінійних двоїстих задач можуть бути застосовані для їхнього розв'язання?

## **ЗМ 2 Економетричні моделі**

### **ТЕМА 9**

## **ПРИНЦИПИ ПОБУДОВИ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ**

### **9.1 Роль економетричних досліджень в економіці**

Економетричне моделювання реальних соціально-економічних процесів і систем зазвичай переслідує одну з двох кінцевих прикладних цілей: прогноз економічних і соціально-економічних показників, що характеризують стан і розвиток аналізованої системи, імітацію різних можливих сценаріїв соціально-економічного розвитку аналізованої системи (різноманітні сценарні розрахунки, ситуаційне моделювання).

При постановці задач економетричного моделювання треба визначити їхній ієрархічний рівень і профіль. Аналізовані задачі можуть відноситися до макро- (країна, міждержавний аналіз), мезо- (регіони усередині країни) і мікро- (підприємства, фірми, родини) рівнів і бути спрямованими на вирішення питань різного профілю інвестиційної, фінансової або соціальної політики, ціноутворення, розподільних стосунків та ін.

Економетрика поєднує сукупність методів і моделей, що дозволяють на базі економічної теорії, економічної статистики і математико-статистичного апарата надавати кількісні вираження якісним залежностям. Основні результати економічної теорії мають якісний характер, а економетрика вносить до них емпіричний зміст. Математична економіка виражає економічні закони у вигляді математичних співвідношень, а економетрика здійснює досліду перевірку цих законів. Економічна статистика дає інформаційне забезпечення досліджуваного процесу у вигляді вихідних (оброблених) статистичних даних і економічних показників, а економетрика, використовуючи традиційні математико-

статистичні і спеціально розроблені методи, аналізує кількісні взаємозв'язки між цими показниками.

Багато базових понять економетрики мають два визначення - економічне і математичне. Подібна подвійність має місце й у формулюванні результатів. Економічна складова економетрики, безумовно, є первинною. Саме економіка визначає постановку задачі і вихідні передумови, а результат, сформований математичною мовою, складає інтерес в тому випадку, якщо вдається його економічна інтерпретація. У той же час велика кількість економетричних результатів має характер математичних стверджень (теорем).

Загальним моментом для будь-якої економетричної моделі є розбивка залежної змінної на дві частини - пояснену і випадкову. Сформулюємо задачу моделювання в самому загальному вигляді: на підставі експериментальних даних визначити пояснену частину і, розглядаючи випадкову складову як випадкову величину, отримати (можливо, після деяких припущень) оцінки параметрів її розподілу. Таким чином, економетрична модель має такий вигляд:

$$\begin{array}{ccccc} \text{Спостережене значення} & & & & \\ \text{залежної змінної} & = & \text{Пояснена частина, залежна від} & + & \text{Випадкова} \\ & & \text{значень змінних, що пояснюють} & & \text{складова} \end{array}$$

Нехай отримано наступний вираз для поясненої частини змінної  $Y$ :

$$y = b + b_1x_1 + b_2x_2.$$

Очевидно, що він дає уявлення про те, як саме формується розглянута економічна змінна  $y$  і можливість виявити вплив на неї кожної з пояснюючих змінних  $x_1$  і  $x_2$ . В цьому випадку пояснювальна змінна  $y$  дорівнює  $b$  при  $x_1=0$  і  $x_2=0$ . За рахунок збільшення  $x_1$  на одну одиницю вона збільшується на  $b_1$ , а за рахунок збільшення  $x_2$  на одну одиницю - зростає на  $b_2$ . Найбільш важливо те, що отриманий вираз дозволяє прогнозувати значення поясненої частини змінної  $Y$ , якщо відомі його параметри  $b$ ,  $b_1$  і  $b_2$ .

## 9.2 Етапи економетричного моделювання

Прийнято виділяти шість основних етапів економетричного моделювання: постановочний, апріорний, етап параметризації, інформаційний, етапи ідентифікації і верифікації моделі.

Зупинимось докладніше на кожному з цих етапів і розглянемо проблеми, пов'язані з їхньою реалізацією.

1-й етап (постановочний). На даному етапі формують ціль дослідження і визначають набір економічних змінних моделі. При виборі економічних змінних необхідно здійснювати теоретичне обґрунтування кожної змінної. При цьому рекомендують, щоб число їх було не дуже великим і, як мінімум, у кілька разів меншим за число спостережень. Пояснюючі змінні не повинні бути пов'язані між собою функціональною або кореляційною залежністю, тому що це може призвести до неможливості оцінки параметрів моделі або до одержання нестійких оцінок, що не мають реального смислу, тобто до явища

мультиколінеарності. Визначальним при включенні до моделі тих або інших змінних є економічний (якісний) аналіз досліджуваного об'єкта.

2-й етап (апріорний). На даному етапі проводять аналіз сутності досліджуваного об'єкта, формування і формалізацію апріорної інформації.

3-й етап (параметризація). На даному етапі здійснюють безпосередньо моделювання, тобто вибір загального вигляду моделі, виявлення зв'язків, які до неї входять. Основна задача, розв'язувана на цьому етапі, - вибір виду функції  $f(x)$ . Досить важливою проблемою на цьому етапі економетричного моделювання є проблема специфікації моделі, зокрема, вираження в математичній формі виявлених зв'язків і співвідношень; визначення складу **екзогенних** (незалежних) і **ендогенних** (залежних) змінних, формулювання вихідних передумов і обмежень моделі. Від того, наскільки вдало вирішена проблема специфікації моделі, у значній мірі залежить успіх всього економетричного моделювання.

4-й етап (інформаційний). На даному етапі здійснюють збір необхідної статистичної інформації - спостережуваних значень економічних змінних

$$(x_{i1}, x_{i2}, \dots, x_{ip}; y_{i1}, y_{i2}, \dots, y_{iq}), i=1, n.$$

Тут можуть бути спостереження, отримані як за участю дослідника, так і без його участі (в умовах активного або пасивного експерименту).

5-й етап (ідентифікація моделі). Тут здійснюють статистичний аналіз моделі і оцінку її параметрів.

6-й етап (верифікація моделі). На цьому етапі проводять перевірку істинності, тобто адекватності моделі. З'ясовують, наскільки вдало вирішені проблеми специфікації, ідентифікації та ідентифікованості моделі, яка точність розрахунків, зроблених на підставі даної моделі, і, в остаточному підсумку, наскільки відповідає побудована модель реальному економічному об'єкту або процесу, який моделюють. Треба помітити, що коли є статистичні дані, які характеризують моделюємий економічний об'єкт у поточний і попередній моменти часу, то для верифікації моделі, побудованої для прогнозу, досить порівняти реальні значення змінних в наступні моменти часу з відповідними їм значеннями, отриманими на основі розглянутої моделі за даними попередніх моментів.

Розглянутий поділ економетричного моделювання на окремі етапи носить певною мірою умовний характер, тому що ці етапи можуть перетинатися, взаємно доповнювати один одного та ін.

### 9.3 Класифікація економетричних моделей

Нехай є  $p$  пояснюючих змінних (факторів)  $x_1, x_2, \dots, x_p$  і залежна змінна  $Y$ . Змінна  $Y$  є випадковою величиною, яка має при заданих значеннях факторів  $x_j$  деякий розподіл. Зазвичай припускають, що умовні розподіли  $Y$  при кожному припустимому значенні факторів  $x_j$  - нормальні.

Пояснюючі змінні  $x_j$  ( $j=1, \dots, p$ ) можуть вважатися як випадковими, так і детермінованими, тобто приймаючими цілком певні значення.

Класична економетрична модель розглядає пояснюючі змінні  $x_j$  як детерміновані.

Пояснена частина  $Y_e$  в кожному разі являє собою функцію від значень факторів - пояснюючих змінних:

$$Y_e = f(x_1, x_2, \dots, x_p).$$

Таким чином, економетрична модель має вигляд

$$Y = f(x_1, x_2, \dots, x_p) + \varepsilon \quad (9.1)$$

Найбільш природним вибором поясненої частини випадкової величини  $Y$  є її середнє значення - умовне математичне сподівання  $M_x(Y)$ , отримане при даному наборі значень пояснюючих змінних  $(x_1, x_2, \dots, x_p)$ .

Рівняння  $M_x(Y) = f(x_1, x_2, \dots, x_p)$  називається **рівнянням регресії**.

При такому природному виборі поясненої частини економетрична модель має вигляд

$$Y = M_x(Y) + \varepsilon, \quad (9.2)$$

де  $\varepsilon$  - випадкова величина, називана **збуренням** або **помилкою**. В курсі математичної статистики рівняння (13.2) називають рівнянням регресійної моделі.

Відзначимо, що економетрична модель не обов'язково є регресійною, тобто пояснена частина не завжди являє собою умовне математичне сподівання залежної змінної. Однією з можливих причин того, що економетрична модель не є регресійною, є наявність систематичних помилок виміру пояснюючих змінних. З математичної точки зору регресійні моделі виявляються істотно більш простим об'єктом, ніж економетрична модель загального типу (13.1).

Щоб одержати досить достовірні та інформативні дані щодо розподілу будь-якої випадкової величини, необхідно мати вибірку її спостережень досить великого обсягу. Така вибірка спостережень залежної змінної  $Y$  і пояснюючих змінних  $x_j$   $i = \overline{1, p}$  є відправною точкою будь-якого економетричного дослідження. Такі вибірки являють собою набори значень  $(x_{i1}, \dots, x_{ip}; y_i)$ , де  $i = \overline{1, n}$ ;  $p$  - кількість пояснюючих змінних;  $n$  - число спостережень.

Як правило, число спостережень  $n$  досить велике й значно перевищує число  $p$  пояснюючих змінних. Проблема, однак, полягає в тому, що спостереження  $y_i$ , розглянуті в різних вибірках як випадкові величини  $Y_i$  і одержувані при різних наборах значень пояснюючих змінних  $x_j$ , мають в загальному випадку різний розподіл. Це означає, що для кожної випадкової величини  $Y_i$  ми маємо всього лише одно спостереження. Зрозуміло, на підставі одного спостереження не можна зробити адекватний висновок щодо розподілу випадкової величини, і потрібні додаткові припущення.

Прийнято виділяти три основних класи моделей, які застосовують для аналізу або прогнозу, це:

- регресійні моделі з одним рівнянням;
- моделі часових рядів;
- системи одночасних рівнянь.



Вибір моделі залежить від вигляду вихідних даних, на підставі яких будують економетричну модель, а також від характеру досліджуваного економічного процесу.

В класичному курсі економетрики розглядають два типи вибірових даних: просторові дані й часові ряди.

**Просторова вибірка.** Прикладом просторових даних є, наприклад, набір відомостей (обсяг виробництва, кількість працівників, дохід та ін.) з різних фірм у той самий момент часу (просторовий зріз). Іншим прикладом можуть бути дані з курсів покупки/продажу наявної валюти в якийсь день в обмінних пунктах міста та ін.

В економіці під просторовою вибіркою розуміють набір показників економічних змінних, отриманий у певний момент часу. Для економетриста, однак, таке визначення не дуже зручне (через неоднозначність поняття «момент часу». Це може бути і день, і тиждень, і рік). Очевидно, про просторову вибірку має сенс говорити в тому випадку, якщо всі спостереження отримані приблизно в незмінних умовах, тобто являють собою набір незалежних вибірових даних з деякої генеральної сукупності.

Будемо називати просторовою вибіркою серію з  $n$  незалежних спостережень  $p$ -мірної випадкової величини  $(x_{i1}, \dots, x_{ip}; y_i)$ . При цьому надалі можна не розглядати  $X_j$  як випадкові величини. В цьому випадку різні випадкові величини  $Y_i$  виявляються між собою незалежними, що спричиняє некорельованість їхніх збурювань, тобто коефіцієнт кореляції між збуреннями  $\varepsilon_i$  і  $\varepsilon_j$

$$r(\varepsilon_i, \varepsilon_j) = 0 \quad \text{при} \quad i=j. \quad (9.3)$$

Умова (13.3) істотно спрощує модель та її статистичний аналіз.

На питання, чи є вибірка серією незалежних спостережень, немає однозначної відповіді. Формально визначення незалежності випадкових величин, як правило, виявляється реально неперевірим. Звичайно за незалежні приймають величини, не зв'язані причинно. Однак на практиці далеко не завжди питання про незалежність виявляється безперечним.

Економетрична модель, побудована на основі просторової вибірки експериментальних даних  $(x_i, y_i)$  є регресійною моделлю з одним рівнянням

$$y_i = f(x_i) + \varepsilon_i, \quad i=1 \dots n, \quad (9.4)$$

де помилки регресії задовольняють умовам

$$M(\varepsilon_i) = 0, \quad (9.5)$$

$$r(\varepsilon_i, \varepsilon_j) = 0, \quad (9.6)$$

$$D(\varepsilon_i) = \sigma_i^2. \quad (9.7)$$

Відносно умови (13.7) можливі два випадки:

а)  $\sigma_i^2 = \sigma_j^2$  при всіх  $i$  та  $j$ . Властивість сталості дисперсій помилок регресії називається **гомоскедастичністю**. В цьому випадку розподіли випадкових величин  $Y_i$  відрізняються тільки значенням математичного сподівання (поясненої частини);

б)  $\sigma_i^2 \neq \sigma_j^2$ . У цьому випадку має місце **гетероскедастичність** моделі. Гетероскедастичність «псує» багато результатів статистичного аналізу і, як правило, вимагає усунення.

В деяких випадках гетероскедастичність моделі очевидна, однак у більшості випадків потрібне застосування методів математичної статистики для прийняття рішення про те, який тип моделі слід розглядати.

Залежно від вигляду функції  $f(x_i)$  моделі ділять на лінійні і нелінійні. Наприклад, можна досліджувати попит на морозиво як функцію від часу, від температури повітря, від середнього рівня доходів або залежність зарплати від віку, статі, рівня освіти, стажу роботи та ін. Область застосування таких моделей, заснованих на просторових даних, навіть лінійних, значно ширша, ніж моделей часових рядів.

**Часовий (динамічний) ряд.** Прикладами часових даних можуть бути щоквартальні дані з інфляції, середньої заробітної плати, національного доходу, грошової емісії за останні роки або, наприклад, щоденний курс долара США на МВБ, ціни ф'ючерсних контрактів на поставку долара США та ін.

Відмінною рисою часових даних є те, що вони природно впорядковані за часом, крім того, спостереження в близькі моменти часу часто бувають залежними.

Часовим (динамічним) рядом називають вибірку спостережень, в якій важливі не тільки самі спостережувані значення випадкових величин, але й порядок їхнього проходження один за одним. Найчастіше впорядкованість обумовлена тим, що експериментальні дані являють собою серію спостережень однієї і тієї ж випадкової величини в послідовні моменти часу. В цьому випадку динамічний ряд називають часовим рядом. При цьому покладають, що тип розподілу спостережуваної випадкової величини залишається тим самим (наприклад, нормальним), але параметри його змінюються залежно від часу.

Моделі часових рядів, як правило, є складніше моделей просторової вибірки, тому що спостереження у випадку часового ряду, загалом кажучи, не є незалежними, а це означає, що помилки регресії можуть корелювати одна з одною, тобто умова (13.3) не виконується. Це значно ускладнює статистичний аналіз моделі.

Відзначимо, що маючи тільки ряд спостережень без розуміння їхньої природи, неможливо визначити, маємо ми справу з просторовою вибіркою або часовим рядом.

До класу моделей часових рядів відносять моделі тренду, сезонності, а також тренду і сезонності.

Модель тренду має вигляд:

$$Y(t) = T(t) + \varepsilon_t, \quad (9.8)$$

де  $T(t)$  - часовий тренд заданого параметричного вигляду (наприклад, лінійний)  $T(t) = a + bt$ ;

$\varepsilon$  - випадкова (стохастична) компонента.

Модель сезонності має вигляд:

$$Y(t) = S(t) + \varepsilon_t, \quad (9.9)$$

де  $S(t)$  - періодична (сезонна) компонента.

Модель тренду і сезонності може бути аддитивною або мультиплікативною моделлю:

$$y(t) = T(t) + S(t) + \varepsilon_t \quad \text{- аддитивна;}$$

$$y(t) = T(t) * S(t) + \varepsilon_t \quad \text{- мультиплікативна,}$$

де  $T(t)$  - часовий тренд заданого параметричного виду;

$S(t)$  - періодична (сезонна) компонента.

До моделей часових рядів належить множина складніших моделей, таких, як моделі адаптивного прогнозу, моделі авторегресії і ковзного середнього та ін. Їхньою загальною рисою є те, що вони пояснюють поведінку часового ряду, виходячи тільки з його попередніх значень. Такі моделі можуть застосовуватися, наприклад, для вивчення й прогнозування обсягу продажів авіаквитків, попиту на морозиво, короткострокового прогнозу процентних ставок та ін.

**Системи одночасних рівнянь.** Моделі, які описуються системами рівнянь, називаються системами одночасних рівнянь. Системи можуть складатися з тотожностей і регресійних рівнянь, кожне з яких може, крім пояснюючих змінних, містити в собі пояснювальні змінні з інших рівнянь системи. Таким чином, тут ми маємо набір пояснювальних змінних, зв'язаних через рівняння системи. Класичним прикладом системи одночасних рівнянь є модель попиту  $Q^d$  і пропозиції  $Q^s$ . Коли попит на товар визначають його ціною  $P$  і доходом споживача  $I$ , пропозиція товару - його ціною  $P$  і досягається рівновага між попитом та пропозицією:

$$Q^s = \alpha_1 + \alpha_2 P + \varepsilon_1 \quad (\text{пропозиція}), \quad (9.10)$$

$$Q^d = \beta_1 + \beta_2 P + \beta_3 I + \varepsilon_2 \quad (\text{попит}), \quad (9.11)$$

$$Q^d = Q^s \quad (\text{рівновага}). \quad (9.12)$$

В цій системі екзогенною (незалежною) змінною є дохід споживача  $I$ , а ендогенними (залежними) – попит (пропозиція) товару  $Q^d = Q^s = Q$  і ціна товару (ціна рівноваги)  $P$ .

В іншій моделі попиту та пропозиції в якості пояснюючої пропозицію змінної може бути не тільки ціна товару  $P$  в даний момент часу  $t$ , але й ціна товару в попередній момент часу  $t-1$ , тобто лагова ендогенна змінна:

$$Q_t^s = \alpha_1 + \alpha_2 P_t + \alpha_3 P_{t-1} + \varepsilon_1. \quad (9.13)$$

Системи одночасних рівнянь вимагають застосування більш складного математичного апарату. Вони можуть використовуватися для моделей странової економіки та ін.

Таким чином, економетрична модель дозволяє пояснити поведінку ендогенних змінних залежно від значень екзогенних і лагових ендогенних змінних, інакше кажучи, залежно від визначених змінних.

Зауважимо, що не всяка економіко-математична модель, що представляє математико-статистичний опис досліджуваного економічного об'єкта, може вважатися економетричною. Вона стає економетричною тільки в тому випадку, якщо буде відбивати цей об'єкт на основі емпіричних (статистичних) даних, що характеризують саме його.

### **Контрольні запитання**

1. З якою метою проводять економетричні дослідження?
2. Поясніть складові загальної економетричної моделі.
3. Охарактеризуйте етапи економетричного моделювання.
4. Поясніть відмінність між ендогенними і екзогенними змінними.
5. У чому полягає ідентифікація і верифікація економетричної моделі?
6. Поясніть терміни «регресія», «умовне математичне сподівання» і «збурення».
7. Які класи моделей використовують для аналізу або прогнозу в економетриці?
8. Якими властивостями повинна володіти регресійна модель з одним рівнянням, побудована на основі просторової вибірки?
9. Поясніть терміни «гомоскедастичність» і «гетероскедастичність».
10. Яку вибірку спостережень називають часовим рядом?
11. Які економетричні моделі належать до систем одночасних рівнянь?

## **ТЕМА 10**

### **МЕТОДИ ПОБУДОВИ ЗАГАЛЬНОЇ ЛІНІЙНОЇ МОДЕЛІ**

#### **10.1 Побудова загальної лінійної моделі**

В практиці економічних досліджень наявні дані не завжди можна вважати вибіркою з багатомірної нормальної сукупності, коли одна з розглянутих змінних не є випадковою або коли лінія регресії явно не пряма та ін. В цих випадках намагаються визначити криву (або поверхню), що дає найкраще наближення до вихідних даних. Відповідні методи наближення одержали назву **регресійного аналізу**.

Методи і моделі регресійного аналізу займають центральне місце в математичному апараті економетрики. Задачами регресійного аналізу є встановлення форми залежності між змінними, оцінка функції регресії, оцінка невідомих значень (прогноз значень) залежної змінної.

В економіці в більшості випадків між змінними величинами існують залежності, в яких кожному значенню однієї змінної відповідає не якийсь певний, а множина можливих значень іншої змінної. Інакше кажучи, кожному значенню однієї змінної відповідає певний (умовний) розподіл іншої змінної. Така залежність одержала назву статистичної (або стохастичної, імовірнісної).

Виникнення поняття статистичного зв'язку обумовлене тим, що залежна змінна піддана впливу ряду неконтрольованих або неврахованих факторів, а крім того тим, що вимір значень змінних неминуче супроводжується деякими випадковими помилками. Прикладом статистичного зв'язку є залежність урожайності від кількості внесених добрив, продуктивності праці на підприємстві від його енергоозброєності та ін.

В силу неоднозначності статистичної залежності між  $Y$  і  $X$  становить інтерес усереднена за  $X$  залежність, тобто закономірність у зміні умовного

математичного сподівання  $M_x(Y)$  (математичного сподівання випадкової змінної  $Y$ , обчисленого в припущенні, що змінна  $X$  прийняла значення  $x$ ) залежно від  $x$ .

Якщо залежність між двома змінними така, що кожному значенню однієї змінної відповідає певне умовне математичне сподівання (середнє значення) іншої, то вона називається **кореляційною**. Інакше кажучи, кореляційною залежністю між двома змінними називають функціональну залежність між значеннями однієї з них і умовним математичним сподіванням іншої.

Кореляційна залежність може бути продана у вигляді

$$M_x(Y) = \varphi(x). \quad (10.1)$$

В регресійному аналізі розглядається однобічна залежність випадкової змінної  $Y$  від однієї (або декількох) не випадкової незалежної змінної  $X$ . Така залежність може виникнути, наприклад, коли при кожному фіксованому значенні  $X$  відповідні значення  $Y$  піддані випадковому розкиду за рахунок дії ряду неконтрольованих факторів. Таку залежність  $Y$  від  $X$  називають **регресійною**. При цьому залежну змінну  $Y$  називають також функцією відгуку, пояснювальною, вихідною, результуючою, ендогенною змінною, результативною ознакою, а незалежну змінну  $X$  - пояснюючою, вхідною, предикторною, екзогенною змінною, фактором, регресором, факторною ознакою.

Рівняння (10.1) називається **рівнянням регресії**, функція  $\varphi(x)$  - **функцією регресії**, а її графік - **лінією регресії**.

Для точного опису рівняння регресії необхідно знати умовний закон розподілу залежної змінної  $Y$  за умови, що змінна  $X$  прийме значення  $x$  ( $X=x$ ). У статистичній практиці таку інформацію одержати, як правило, не вдається, тому що звичайно дослідник розташовує лише вибіркою пар значень  $(x_i, y_i)$  обмеженого обсягу  $n$ . У цьому випадку мова може йти про оцінку (наближений вираз, апроксимацію) за вибіркою функції регресії. Такою оцінкою є вибіркова лінія регресії:

$$\hat{y} = \hat{\varphi}(x, b_0, b_1, \dots, b_p), \quad (10.2)$$

де  $y$  - умовна (групові) середня змінної  $Y$  при фіксованому значенні змінної  $X=x$ ;

$b_0, b_1, \dots, b_p$  - параметри кривої.

Рівняння (10.2) називають вибірковим рівнянням регресії.

При правильно визначеній апроксимуючій функції  $\varphi(x, b_0, b_1, \dots, b_p)$  із збільшенням обсягу вибірки ( $n \rightarrow \infty$ ) вона буде збігатися за імовірністю до функції регресії  $\varphi(x)$ .

Модель парної регресії називають **простою** (з двома змінними) **економетричною моделлю**. Її застосовують, якщо завдання полягає у визначенні або в оцінці залежності пояснювальної змінної  $Y$  від одного фактору (пояснюючої змінної)  $X$ . При цьому вихідні дані являють собою два набори значень (вектори)  $x = (x_1, \dots, x_n)$  і  $y = (y_1, \dots, y_n)$ . Метою є одержання аналітичної залежності пояснювальної змінної  $Y$  від пояснюючої змінної  $x$ , яка щонайкраще відповідає вихідним даним.

Таким чином, парна регресія являє собою регресію між двома змінними –  $y$  та  $x$ , тобто модель виду:

$$y = \hat{f}(x), \quad (10.3)$$

де  $y$  - залежна змінна (результативна ознака);

$x$  - незалежна, або пояснююча, змінна (ознака-фактор).

Знак « $\wedge$ » означає, що між змінними  $x$  та  $y$  немає строгої функціональної залежності, тому практично в кожному окремому випадку величина  $y$  складається із двох доданків:

$$y = \hat{y}_x + \varepsilon, \quad (10.4)$$

де  $y$  – фактичне значення результативної ознаки;

$\hat{y}_x$  – теоретичне значення результативної ознаки, знайдене виходячи з рівняння регресії;

$\varepsilon$  – випадкова величина, що характеризує відхилення реального значення результативної ознаки від теоретичного, знайденого за рівнянням регресії.

Випадкову величину  $\varepsilon$  називають також **збуренням**. Вона включає вплив не врахованих у моделі факторів, випадкових помилок і особливостей виміру. Її присутність в моделі породжено трьома джерелами: специфікацією моделі, вибіркоvim характером вихідних даних і особливостями виміру змінних.

В регресійному аналізі розглядають односторонню залежність випадкової змінної  $Y$  від однієї (або декількох) неслучайної незалежної змінної  $X$ .

Від правильно обраної специфікації моделі залежить величина випадкових помилок: вони тим менші, чим менше теоретичні значення результативної ознаки  $\hat{y}_x$  відрізняються від фактичних даних  $y$ .

До помилок специфікації відносять неправильний вибір тієї або іншої математичної функції для  $\hat{y}_x$  і не врахування у рівнянні регресії будь-якого істотного фактору, тобто використання парної регресії замість множинної.

Поряд з помилками специфікації можуть мати місце помилки вибірки, які звичайно обумовлені неоднорідністю даних у вихідній статистичній сукупності. Це, як правило, має місце при вивченні економічних процесів. Якщо сукупність неоднорідна, то рівняння регресії не має практичного смислу. Для одержання доброго результату звичайно виключають із сукупності одиниці з аномальними значеннями досліджуваних ознак.

Використання часової інформації також являє собою вибірку з усієї множини хронологічних дат. Змінивши часовий інтервал, можна одержати інші результати.

Найбільшою небезпекою в практичному використанні методів регресії є помилки виміру. Якщо помилки специфікації можна зменшити, змінюючи форму моделі (вигляд математичного виразу), а помилки вибірки - збільшуючи обсяг вихідних даних, то помилки виміру практично зводять нанівець усі зусилля з кількісної оцінки зв'язку між ознаками.

Особливо велику роль мають помилки виміру при дослідженні на макрорівні. Так, у дослідженнях попиту і споживання в якості пояснюючої змінної широко використовують «дохід населення». Разом з тим, статистичний вимір величини доходу пов'язано з рядом труднощів і не позбавлено можливих помилок, наприклад, в результаті наявності схованих доходів.

Припускаючи, що помилки виміру зведені до мінімуму, основну увагу в економетричних дослідженнях приділяють помилкам специфікації моделі.

В парній регресії вибір вигляду математичної функції  $\hat{y}_x = f(x)$  може бути здійснений за трьома методами:

- графічним;
- аналітичним, тобто виходячи з теорії досліджуваного взаємозв'язку;
- експериментальним.

При вивченні залежності між двома ознаками графічний метод підбору виду рівняння регресії є досить наочним. Він заснований на побудові **поля кореляції**. Основні типи кривих, які використовують при кількісній оцінці зв'язків, зображені на рисунку 10.1.

Значну зацікавленість викликає аналітичний метод вибору типу рівняння регресії. Він заснований на вивченні матеріальної природи зв'язку досліджуваних ознак.

При обробці інформації на комп'ютері вибір виду рівняння регресії звичайно здійснюють за експериментальним методом, тобто шляхом порівняння величини залишкової дисперсії  $\sigma_{\text{зал}}^2$ , яку розраховують при різних моделях.

Якщо рівняння регресії охоплює всі точки кореляційного поля, що можливо тільки при функціональному зв'язку, коли всі точки лежать на лінії регресії  $\hat{y}_x = f(x)$ , то фактичні значення результативної ознаки збігаються з теоретичними  $y = \hat{y}_x$ , тобто вони повністю обумовлені впливом фактора  $x$ . В цьому випадку залишкова дисперсія  $\sigma_{\text{зал}}^2 = 0$ .

В практичних дослідженнях, як правило, має місце деяке розсіювання точок кореляційного поля щодо лінії регресії. Воно обумовлене впливом інших факторів, які не враховані у рівнянні регресії. Іншими словами, мають місце відхилення фактичних даних від теоретичних  $(y - \hat{y}_x)$ . Величина цих відхилень лежить в основі розрахунку залишкової дисперсії:

$$\sigma_{\text{зал}}^2 = \frac{1}{n} \sum_{i=1}^n (y - \hat{y}_x)^2. \quad (10.5)$$

Чим меншою є величина залишкової дисперсії, тим меншим є вплив факторів, які не враховуються у рівнянні регресії, і тим краще рівняння регресії підходить до вихідних даних.

Вважається, що число спостережень має в 7-8 разів перевищувати число параметрів при змінній  $x$ , що розраховуються. Це означає, що шукати лінійну регресію, маючи менш за 7 спостережень, взагалі не має рації. Якщо вигляд функції ускладнюється, то необхідно збільшити обсяг спостережень, тому що

кожний параметр при  $x$  повинен розраховуватися хоча б за 7 спостереженнями. Виходить, якщо ми вибираємо криву другого ступеня (параболу)  $\hat{y}_x = a + bx + cx^2$ , то потрібен обсяг інформації вже не менший за 14 спостережень.

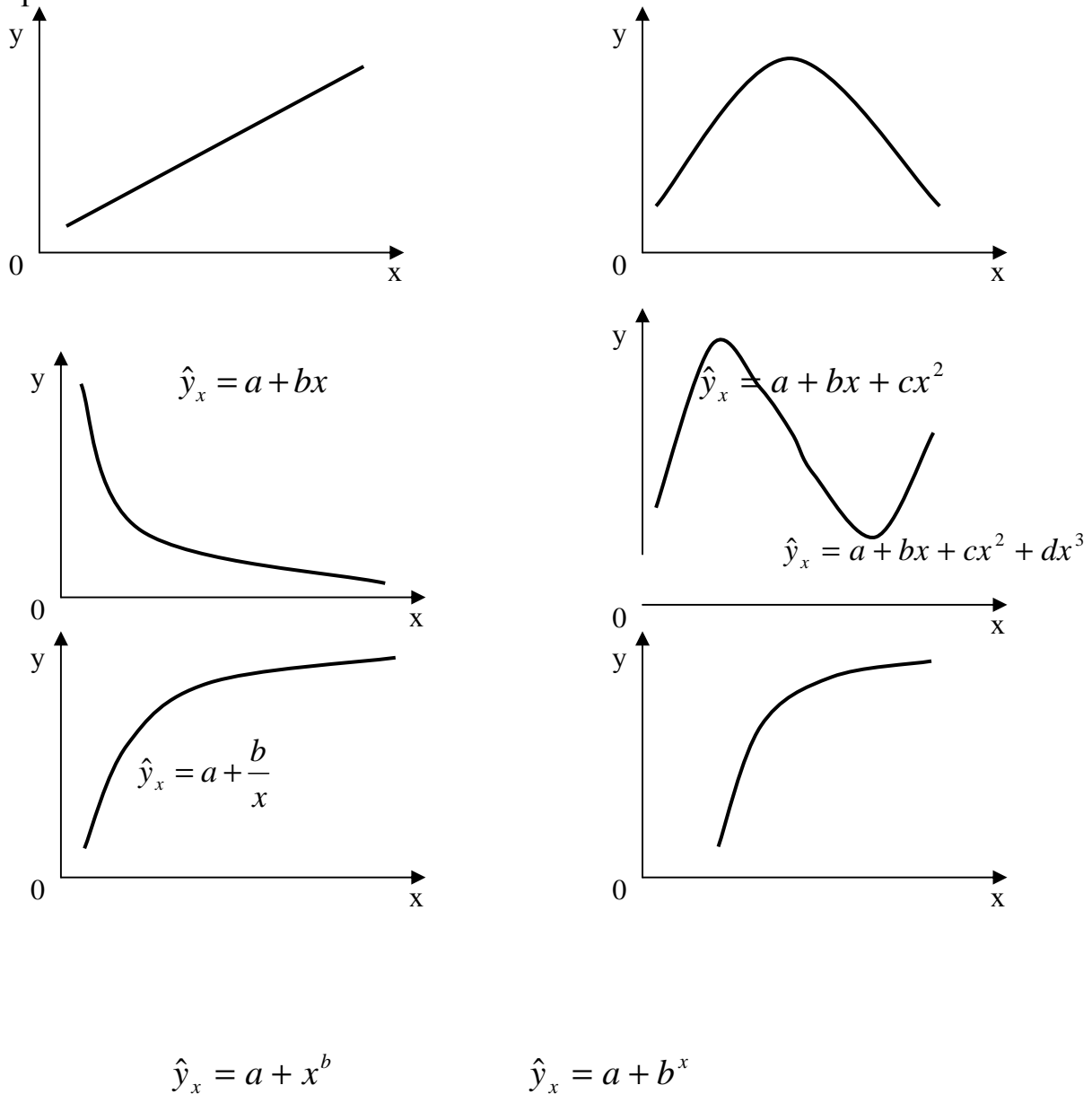


Рисунок 10.1 - Основні типи кривих, які використовують при кількісній оцінці зв'язків між двома змінними

## 10.2 Лінійна модель парної регресії. Сутність методу найменших квадратів

Найкраще вивчені лінійні регресійні моделі, що задовольняють умовам (10.5), (10.6) і властивості сталості дисперсії помилок регресії, - їх називають **класичними моделями**.

Помітимо, що умовам класичної регресійної моделі задовольняють і гомоскедастична модель просторової вибірки, і модель часового ряду,



спостереження якого не корелюють, а дисперсії постійні. З математичної точки зору вони дійсно нерозрізнювані (хоча можуть значно розрізнятися економічні інтерпретації отриманих математичних результатів). Нехай визначений характер експериментальних даних і виділений певний набір пояснюючих змінних. Для того щоб знайти пояснену частину, тобто величину  $M_x(Y)$ , потрібне знання умовних розподілів випадкової величини  $Y$ . На практиці це майже ніколи не має місця, тому точне знаходження поясненої частини неможливе. В таких випадках застосовують стандартну процедуру згладжування експериментальних даних. Ця процедура складається з двох етапів:

- визначають параметричне сімейство, до якого належить шукана функція  $M_x(Y)$  (розглянута як функція від значень пояснюючих змінних  $X$ ). Це може бути множина лінійних функцій, степеневих функцій та ін. (рис. 10.1);

- визначають оцінки параметрів цієї функції за допомогою одного з методів математичної статистики.

Формально ніяких способів вибору параметричного сімейства не існує. Однак у переважній більшості випадків економетричні моделі вибираються лінійними. Окрім цілком очевидної переваги лінійної моделі - її відносної простоти, - для такого вибору є, принаймні, дві істотні причини. Перша причина: якщо випадкова величина  $(X, Y)$  має спільний нормальний розподіл, тоді, як відомо, рівняння регресії лінійні. Припущення про нормальний розподіл є цілком природним і в ряді випадків може бути обґрунтованим за допомогою граничних теорем теорії імовірностей.

В інших випадках самі величини  $Y$  або  $X$  можуть не мати нормального розподілу, але деякі функції від них розподілені нормально. Наприклад, відомо, що логарифм доходів населення - нормально розподілена випадкова величина. Цілком природно, наприклад, вважати нормально розподіленою випадковою величиною пробіг автомобіля. Часто гіпотезу про нормальний розподіл приймають в багатьох випадках, коли немає явного їй протиріччя.

Друга причина, за якою лінійна регресійна модель виявляється переважніше інших, - це менший ризик значної помилки прогнозу. Лінійна регресія знаходить широке застосування в економетриці через чітку економічну інтерпретацію її параметрів.

Лінійна регресія зводиться до знаходження рівняння виду

$$\hat{y}_x = a + bx \text{ або } y = a + bx + \varepsilon. \quad (10.6)$$

Рівняння виду  $\hat{y}_x = a + bx$  дозволяє за заданим значенням фактору  $x$  знаходити теоретичні значення результативної ознаки, підставляючи до нього фактичні значення фактора  $x$ .

Побудова лінійної регресії зводиться до оцінки її параметрів -  $a$  і  $b$ . Класичний підхід до оцінювання параметрів лінійної регресії заснований на **методі найменших квадратів** (МНК), що відповідно до теореми Гауса-Маркова дає найкращі оцінки цих параметрів.

**Теорема Гауса-Маркова.** Якщо регресійна модель (10.6) має наступні властивості:

- збурення  $\varepsilon$  (або залежна змінна  $y_i$ ) є величиною випадковою, а пояснююча змінна  $x_i$  - величиною не випадковою;
  - математичне сподівання збурення  $\varepsilon$  дорівнює нулю;
  - дисперсія збурення  $\varepsilon$  (або залежної змінної  $y_i$ ) постійна для будь-якого  $i$  (або  $D(y_i) = \sigma^2$ ) - умова гомоскедастичності або рівномірності збурення (залежної змінної);
  - збурення  $\varepsilon_i$  і  $\varepsilon_j$  (або змінні  $y_i$  і  $y_j$ ) не корельовані;
  - збурення  $\varepsilon_i$  (або залежна змінна  $y_i$ ) є нормально розподіленою випадковою величиною,
- то оцінки параметрів лінійної регресії мають найменшу дисперсію в класі всіх лінійних незміщених оцінок.

Отже, оцінки параметрів лінійної регресії є **ефективними** оцінками.

МНК дозволяє одержати такі оцінки параметрів  $a$  і  $b$ , при яких сума квадратів відхилень фактичних значень результативної ознаки  $y$  від теоретичних  $\hat{y}_x$  мінімальна:

$$\sum_{i=1}^n (y_i - \hat{y}_{xi})^2 = \sum_{i=1}^n \varepsilon_i^2 \rightarrow \min. \quad (10.7)$$

Тобто з усієї множини ліній лінія регресії на графіку вибирається так, щоб сума квадратів відстаней за вертикаллю між статистичними точками і цією лінією була б мінімальною (рис. 10.2).

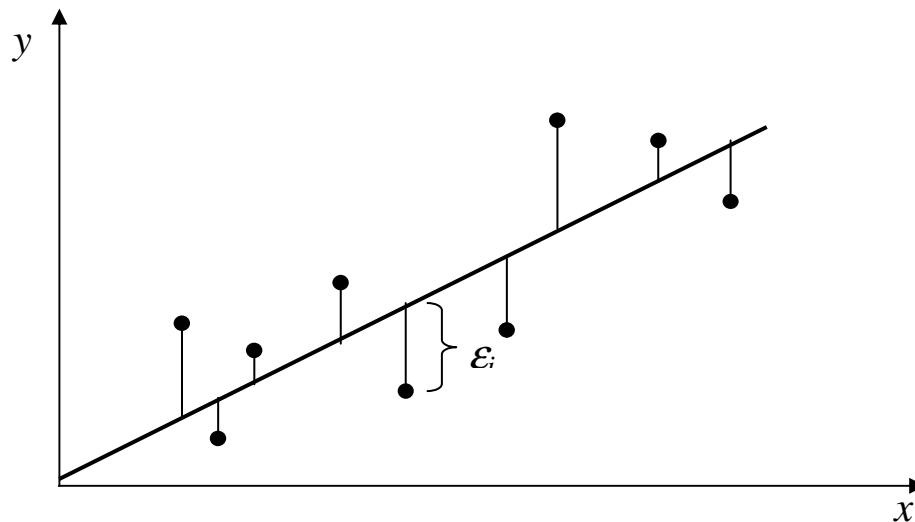


Рисунок 10.2 - Лінія регресії з мінімальною дисперсією залишків

Як відомо з курсу математичного аналізу, щоб знайти мінімум функції (10.7), треба обчислити часткові похідні за кожним з параметрів  $a$  і  $b$  і дорівняти їх до нуля. Позначимо  $\sum_{i=1}^n \varepsilon_i^2$  через  $S(a, b)$ , тоді:

$$S(a, b) = \sum_{i=1}^n (y_i - a - bx_i)^2,$$

$$\begin{cases} \frac{\partial S}{\partial a} = -2 \sum_{i=1}^n (y - a - bx) = 0, \\ \frac{\partial S}{\partial b} = -2 \sum_{i=1}^n (y - a - bx)x = 0. \end{cases} \quad (10.8)$$

Після нескладних перетворень одержимо наступну систему лінійних рівнянь для оцінки параметрів  $a$  і  $b$ :

$$\begin{cases} b * \sum x_i^2 + a * \sum x_i = \sum x_i y_i, \\ b * \sum x_i + na = \sum y_i. \end{cases} \quad (10.9)$$

Вирішуючи систему рівнянь (10.9), знайдемо шукані оцінки параметрів  $a$  і  $b$ . Можна скористатися наступними готовими формулами, які впливають безпосередньо з розв'язання системи (10.9):

$$a = \bar{y} - b\bar{x}, \quad b = \frac{\text{cov}(x, y)}{\sigma_x^2}, \quad (10.10)$$

де  $\text{cov}(x, y) = \overline{yx} - \bar{y} * \bar{x}$  – коваріація ознак  $x$  і  $y$ ;

$\sigma_x^2 = \overline{x^2} - \bar{x}^2$  – дисперсія ознаки  $x$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y, \quad \overline{yx} = \frac{1}{n} \sum_{i=1}^n yx, \quad \overline{x^2} = \frac{1}{n} \sum_{i=1}^n x^2.$$

Коваріація – числова характеристика спільного розподілу двох випадкових величин, яка дорівнює математичному сподіванню добутку відхилень цих випадкових величин від їхніх математичних сподівань. Дисперсія – характеристика випадкової величини, що визначається як математичне сподівання квадрата відхилення випадкової величини від її математичного сподівання. Математичне сподівання – сума добутків значень випадкової величини та відповідних імовірностей.

Параметр  $b$  називають **коефіцієнтом регресії**. Його величина показує середню зміну результату зі зміною фактора на одну одиницю. Можливість чіткої економічної інтерпретації коефіцієнта регресії зробила лінійне рівняння регресії досить розповсюдженим в економетричних дослідженнях.

Формально  $a$  – значення  $y$  при  $x=0$ . Якщо фактор-ознака  $x$  не може мати нульового значення, то вищевказане трактування вільного члена  $a$  не має значення, тобто параметр  $a$  може не мати економічного смислу.

Рівняння регресії завжди доповнюється показником тісноти зв'язку результативної ознаки  $Y$  і фактора-ознаки  $X$ . При використанні лінійної регресії як такий показник виступає лінійний **коефіцієнт кореляції**  $r_{xy}$ , який можна розрахувати за наступними формулами:

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}. \quad (10.11)$$

Лінійний коефіцієнт кореляції знаходиться в межах:  $-1 \leq r_{xy} \leq 1$ . Чим ближче абсолютне значення  $r_{xy}$  до одиниці, тим сильніше лінійний зв'язок між факторами (при  $r_{xy} = \pm 1$  має місце строго функціональна залежність). Але

необхідно мати на увазі, що близькість абсолютної величини лінійного коефіцієнта кореляції до нуля ще не означає відсутності зв'язку між ознаками. При іншій (нелінійній) специфікації моделі зв'язок між ознаками може виявитися досить тісним.

Для оцінки якості підбору лінійної функції розраховують квадрат лінійного коефіцієнта кореляції  $r_{xy}^2$ , який називають **коефіцієнтом детермінації**. Коефіцієнт детермінації характеризує частку дисперсії результативної ознаки  $y$ , що пояснюється регресією, в загальній дисперсії результативної ознаки:

$$r_{xy}^2 = 1 - \frac{\sigma_{\text{зал}}^2}{\sigma_y^2}, \quad (10.12)$$

$$\text{де } \sigma_{\text{зал}}^2 = \frac{1}{n} \sum_{i=1}^n (y - \hat{y}_x)^2, \quad \sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y - \bar{y})^2 = \overline{y^2} - \bar{y}^2.$$

Відповідно величина  $1 - r_{xy}^2$  характеризує частку дисперсії  $y$ , викликану впливом інших, не врахованих в моделі, факторів.

Після того як рівняння лінійної регресії знайдене, проводять оцінку значущості як рівняння в цілому, так і окремих його параметрів.

### 10.3 Оцінка значущості рівняння лінійної регресії та перевірка моделі на адекватність за критеріями Стюдента і Фішера

Перевірити значущість рівняння регресії – означає встановити, чи відповідає математична модель, що виражає залежність між змінними, експериментальним даним і чи достатньо включених до рівняння пояснюючих змінних (однієї або декількох) для опису залежної змінної.

Щоб мати загальне судження про якість моделі, визначають середню помилку апроксимації:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y - \hat{y}_x}{y} \right| * 100\%. \quad (10.13)$$

Середня помилка апроксимації не повинна перевищувати 8-10%.

Оцінка значущості рівняння регресії в цілому здійснюється за F-критерієм Фішера, розрахунку якого передують дисперсійний аналіз. В математичній статистиці дисперсійний аналіз розглядають як самостійний інструмент статистичного аналізу. В економетриці його застосовують як допоміжний засіб для вивчення якості регресійної моделі.

Відповідно до основної ідеї дисперсійного аналізу, загальну суму квадратів відхилень змінної  $y$  від середнього значення  $\bar{y}$  розкладають на дві частини – пояснену і непояснену:

$$\sum_{i=1}^n (y - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_x - \bar{y})^2 + \sum_{i=1}^n (y - \hat{y}_x)^2, \quad (10.14)$$

де  $\sum_{i=1}^n (y - \bar{y})^2$  – загальна сума квадратів відхилень;

$\sum_{i=1}^n (\hat{y}_x - \bar{y})^2$  – сума квадратів відхилень, пояснена регресією (або факторна сума квадратів відхилень);

$\sum_{i=1}^n (y - \hat{y}_x)^2$  – залишкова сума квадратів відхилень, що характеризує вплив неврахованих у моделі факторів.

Схема дисперсійного аналізу має вигляд, наданий таблицею 10.1 ( $n$  – число спостережень,  $m$  – число параметрів при змінній  $x$ ).

Таблиця 10.1 - Схема дисперсійного аналізу

Компоненти дисперсії	Сума квадратів	Число ступенів свободи	Дисперсія на один ступінь свободи
Загальна	$\sum_{i=1}^n (y - \bar{y})^2$	$n-1$	$S_{\text{общ}}^2 = \frac{\sum_{i=1}^n (y - \bar{y})^2}{n-1}$
Факторна	$\sum_{i=1}^n (\hat{y}_x - \bar{y})^2$	$m$	$S_{\text{факт}}^2 = \frac{\sum_{i=1}^n (\hat{y}_x - \bar{y})^2}{m}$
Залишкова	$\sum_{i=1}^n (y - \hat{y}_x)^2$	$n-m-1$	$S_{\text{зал}}^2 = \frac{\sum_{i=1}^n (y - \hat{y}_x)^2}{n-m-1}$

Визначення дисперсії на один ступінь свободи приводить дисперсії до порівнянного виду. Зіставляючи факторну і залишкову дисперсії, що розраховані на один ступінь свободи, одержимо величину  $F$ -критерію Фішера:

$$F = \frac{S_{\text{факт}}^2}{S_{\text{зал}}^2}. \quad (10.15)$$

Фактичне значення  $F$ -критерію Фішера (10.15) порівнюють з табличним значенням  $F_{\text{табл}}(\alpha, k_1, k_2)$  при рівні значущості  $\alpha$  і ступенях свободи  $k_1=m$  і  $k_2=n-m-1$ . При цьому якщо фактичне значення  $F$ -критерію перевищує табличне, то визнають статистичну значущість рівняння в цілому.

Для парної лінійної регресії  $m=1$ , тому

$$F = \frac{S_{\text{факт}}^2}{S_{\text{зал}}^2} = \frac{\sum (\hat{y}_x - \bar{y})^2}{\sum (y - \hat{y}_x)^2} * (n-2). \quad (10.16)$$

Величина  $F$ -критерію пов'язана з коефіцієнтом детермінації  $r_{xy}^2$ , і її можна розрахувати за наступною формулою:

$$F = \frac{r_{xy}^2}{1-r_{xy}^2} * (n-2). \quad (10.17)$$

В парній лінійній регресії оцінюють значущість не тільки рівняння в цілому, але й окремих його параметрів. З цією метою для кожного з параметрів визначають його стандартну помилку:  $m_a$  і  $m_b$ .

Стандартну помилку коефіцієнта регресії визначають за формулою:

$$m_b = \sqrt{\frac{S_{\text{зал}}^2}{\sum (x - \bar{x})^2}} = \frac{S_{\text{зал}}}{\sigma_x \cdot \sqrt{n}}, \quad (10.18)$$

де  $S_{\text{зал}}^2 = \frac{\sum (y - \hat{y}_x)^2}{n - 2}$  – залишкова дисперсія на один ступінь свободи.

Величину стандартної помилки разом з  $t$ -розподілом Стьюдента при  $n-2$  ступенях свободи застосовують для перевірки значущості коефіцієнта регресії і для розрахунку його довірчого інтервалу.

Для оцінки значущості коефіцієнта регресії його величину порівнюють з його стандартною помилкою, тобто визначають фактичне значення  $t$ -критерію

Стьюдента:  $t_b = \frac{b}{m_b}$ , яке потім порівнюють з табличним значенням за певним

рівнем значущості  $\alpha$  і числом ступенів свободи  $n-2$ . Довірчий інтервал для коефіцієнта регресії визначають як  $b \pm t_{\text{табл}} m_b$ . Оскільки знак коефіцієнта регресії вказує на зростання результативної ознаки у при збільшенні ознаки-фактора  $x$  ( $b > 0$ ), зменшення результативної ознаки при збільшенні ознаки-фактора ( $b < 0$ ) або його незалежність від незалежної змінної ( $b = 0$ ), то межі довірчого інтервалу для коефіцієнта регресії не повинні містити суперечливих результатів (наприклад,  $-1,5 \leq b \leq 0,8$ ). Такого роду запис вказує, що дійсне значення коефіцієнта регресії одночасно містить додатні і від'ємні величини і навіть нуль, чого не може бути.

Стандартну помилку параметра  $a$  визначають за формулою:

$$m_a = \sqrt{S_{\text{зал}}^2 \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = S_{\text{зал}} \frac{\sqrt{\sum x^2}}{\sigma_x n}. \quad (10.19)$$

Процедура оцінювання значущості даного параметра не відрізняється від розглянутої вище для коефіцієнта регресії. Обчислюють  $t$ -критерій:  $t_a = \frac{a}{m_a}$ ,

його величину порівнюють з табличним значенням при  $n-2$  ступенях свободи.

Значущість лінійного коефіцієнта кореляції перевіряють на основі величини помилки коефіцієнта кореляції  $m_r$ :

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}. \quad (10.20)$$

Фактичне значення  $t$ -критерію Стьюдента визначається як  $t_r = \frac{r}{m_r}$ .

Існує зв'язок між  $t$ -критерієм Стьюдента і  $F$ -критерієм Фішера:

$$t_b = t_r = \sqrt{F}. \quad (10.21)$$

В прогнозних розрахунках за рівнянням регресії визначають прогнозоване  $\hat{y}_p$  значення як точковий прогноз  $\hat{y}_x$  при  $x_p = x_k$ , тобто шляхом

підстановки до рівняння регресії  $\hat{y}_x = a + bx$  відповідного значення  $x$ . Він доповнюється розрахунком стандартної помилки  $\hat{y}_p$ , тобто  $M(\hat{y}_p)$ , і відповідною інтервальною оцінкою прогнозного значення  $\hat{y}_p$ :

$$\hat{y}_p - \Delta_{\hat{y}_p} \leq \hat{y}_p \leq \hat{y}_p + \Delta_{\hat{y}_p},$$

де  $\Delta_{\hat{y}_p} = m_{\hat{y}_p} * t_{табл}$ , а  $m_{\hat{y}_p}$  – середня помилка прогнозованого індивідуального значення:

$$m_{\hat{y}_p} = S_{зал} \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2}}. \quad (10.22)$$

#### 10.4 Лінійна модель множинної регресії

Економічні явища, як правило, визначаються великим числом одночасно і сукупно діючих факторів. У зв'язку з цим часто виникає задача дослідження залежності однієї залежної змінної  $Y$  від кількох пояснюючих змінних  $X_1, X_2, \dots, X_n$ . Цю задачу вирішують за допомогою **множинного регресійного аналізу**. Парна регресія може дати добрий результат при моделюванні, якщо впливом інших факторів, можна зневажити. Якщо ж цим впливом зневажити не можна, то слід спробувати виявити вплив інших факторів, увівши їх до моделі, тобто побудувати рівняння множинної регресії

$$\hat{y} = f(x_1, x_2, \dots, x_m), \quad (10.23)$$

де  $y$  – залежна змінна (результативна ознака);

$x_i$  – незалежні, або пояснюючі, змінні (ознаки-фактори).

Множинну регресію широко використовують в розв'язанні проблем попиту, прибутковості акцій, при вивченні функції витрат виробництва, у макроекономічних розрахунках і цілому ряді інших питань економетрики. Основна мета множинної регресії – побудувати модель із великим числом факторів, визначивши при цьому вплив кожного з них окремо, а також сукупний їхній вплив на показник, що моделюється.

У зв'язку з чіткою інтерпретацією параметрів найбільш широко використовують лінійну функцію. У лінійній множинній регресії  $\hat{y}_x = a + b_1x_1 + b_2x_2 + \dots + b_mx_m$  параметри при  $x$  називаються коефіцієнтами «чистої» регресії. Вони характеризують середню зміну результативної ознаки із зміною відповідного фактору на одиницю при незмінному значенні інших факторів, зафіксованих на середньому рівні.

Розглянемо лінійну модель множинної регресії

$$y = a + b_1x_1 + b_2x_2 + \dots + b_mx_m + \varepsilon. \quad (10.24)$$

Класичний підхід до оцінювання параметрів лінійної моделі множинної регресії заснований на методі найменших квадратів (МНК). Теорема Гауса-Маркова, що розглянута для парної регресійної моделі, виявляється вірною і для моделі множинної регресії. Тобто МНК дає найбільш ефективні (тобто такі,

що володіють найменшою дисперсією в класі лінійних незміщених оцінок) оцінки параметрів регресійної моделі.

Нагадаємо, що МНК дозволяє одержати такі оцінки параметрів, за якими сума квадратів відхилень фактичних значень результативної ознаки у від розрахункових  $\hat{y}_x$  мінімальна:

$$\sum_i \left( y_i - \hat{y}_{x_i} \right)^2 \rightarrow \min. \quad (10.25)$$

З виразу (10.25) дістанемо функцію  $m+1$  аргументу:

$$S(a, b_1, b_2, \dots, b_m) = \sum_{i=1}^n (y - a - b_1 x_1 - b_2 x_2 - \dots - b_m x_m)^2.$$

Визначивши часткові похідні першого порядку, дорівнюємо їх до нуля:

$$\left\{ \begin{aligned} \frac{\partial S}{\partial a} &= -2 \sum_{i=1}^n (y - a - b_1 x_1 - b_2 x_2 - \dots - b_m x_m) = 0, \\ \frac{\partial S}{\partial b_1} &= -2 \sum_{i=1}^n (y - a - b_1 x_1 - b_2 x_2 - \dots - b_m x_m) x_1 = 0, \\ &\dots \dots \dots \\ \frac{\partial S}{\partial b_m} &= -2 \sum_{i=1}^n (y - a - b_1 x_1 - b_2 x_2 - \dots - b_m x_m) x_m = 0. \end{aligned} \right.$$

Після елементарних перетворень приходимо до системи лінійних нормальних рівнянь для знаходження параметрів лінійного рівняння множинної регресії (10.24):

$$\left\{ \begin{array}{l} na + b_1 \sum_{i=1}^n x_{i1} + b_2 \sum_{i=1}^n x_{i2} + \dots + b_m \sum_{i=1}^n x_{im} = \sum_{i=1}^n y_i, \\ a \sum_{i=1}^n x_{i1} + b_1 \sum_{i=1}^n x_{i1}^2 + b_2 \sum_{i=1}^n x_{i1}x_{i2} + \dots + b_m \sum_{i=1}^n x_{i1}x_{im} = \sum_{i=1}^n x_{i1}y_i, \\ \dots\dots\dots \\ a \sum_{i=1}^n x_{im} + b_1 \sum_{i=1}^n x_{i1}x_{im} + b_2 \sum_{i=1}^n x_{i2}x_{im} + \dots + b_m \sum_{i=1}^n x_{im}^2 = \sum_{i=1}^n x_{im}y_i. \end{array} \right. \quad (10.26)$$

Для двофакторної моделі дана система матиме вигляд:

$$\begin{cases} na + b_1 \sum_{i=1}^n x_{i1} + b_2 \sum_{i=1}^n x_{i2} = \sum_{i=1}^n y_i, \\ a \sum_{i=1}^n x_{i1} + b_1 \sum_{i=1}^n x_{i1}^2 + b_2 \sum_{i=1}^n x_{i1} x_{i2} = \sum_{i=1}^n x_{i1} y_i, \\ a \sum_{i=1}^n x_{i2} + b_1 \sum_{i=1}^n x_{i1} x_{i2} + b_2 \sum_{i=1}^n x_{i2}^2 = \sum_{i=1}^n x_{i2} y_i. \end{cases} \quad (10.27)$$

На практиці часто виникає необхідність у порівнянні впливу на результативну ознаку різних пояснюючих змінних. Для цього необхідно, щоб ці пояснюючі змінні були приведені до однакових одиниць виміру. В таких випадках використовують стандартизовані коефіцієнти регресії:



$$t_y = \frac{\bar{y} - y}{\sigma_y}. \quad (10.28)$$

Метод найменших квадратів можна застосувати і до рівняння множинної регресії в стандартизованому вигляді:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_m t_{x_m} + \varepsilon, \quad (10.29)$$

де  $t_y, t_{x_1}, \dots, t_{x_m}$  – стандартизовані змінні:  $t_y = \frac{y - \bar{y}}{\sigma_y}$ ,  $t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$ , для яких

середнє значення дорівнює нулю  $\bar{t}_y = \bar{t}_{x_i} = 0$ , а середнє квадратичне відхилення дорівнює одиниці:  $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$ ;  $\beta_i$  – стандартизовані коефіцієнти регресії.

Стандартизовані коефіцієнти регресії показують, на скільки одиниць зміниться в середньому результат, якщо відповідний фактор  $x_i$  зміниться на одну одиницю при незмінному середньому рівні інших факторів. В силу того, що всі змінні задані як центровані й нормовані, стандартизовані коефіцієнти регресії  $\beta_i$  можна порівнювати один з одним. Порівнюючи їх один з одним, можна ранжирувати фактори за силою їхнього впливу на результативну ознаку. В цьому полягає основна перевага стандартизованих коефіцієнтів регресії на відміну від коефіцієнтів «чистої» регресії, які непорівнянні один з одним.

Застосовуючи МНК до рівняння множинної регресії в стандартизованому вигляді, одержимо систему нормальних рівнянь

[illegible]

де  $r_{yx_i}$  і  $r_{x_j x_i}$  – коефіцієнти парної і міжфакторної кореляції.

Коефіцієнти «чистої» регресії  $b_i$  пов'язані із стандартизованими коефіцієнтами регресії  $\beta_i$  у такий спосіб:

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}. \quad (10.31)$$

Тому можна переходити від рівняння регресії в стандартизованому масштабі (10.29) до рівняння регресії в натуральному масштабі змінних (10.24), при цьому параметр  $a$  визначається як  $a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - \dots - b_m \bar{x}_m$ .

Зміст стандартизованих коефіцієнтів регресії дозволяє використовувати їх при відсіванні факторів – з моделі виключають фактори з найменшим значенням  $\beta_i$ .

На основі лінійного рівняння множинної регресії (10.24) можуть бути знайдені часткові рівняння регресії:

$$(10.32)$$

[illegible][illegible]
$$\left\{ \begin{array}{l} A_1 = a + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \dots + b_m \bar{x}_m, \\ A_2 = a + b_1 \bar{x}_1 + b_3 \bar{x}_3 + \dots + b_m \bar{x}_m, \\ \dots\dots\dots \\ A_m = a + b_1 \bar{x}_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \dots + b_{m-1} x_{m-1}. \end{array} \right.$$
$$E_{y_{x_i}} = b_i \frac{x_i}{\hat{y}_{x_1 x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_m}}, \quad (10.34)$$
$$\hat{y}_{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m} - \text{часткове рівняння регресії.}$$

106

$$\overline{E}_i = b_i \frac{\overline{x}_i}{\overline{y}_{x_i}}, \quad (10.35)$$

які показують, на скільки відсотків у середньому зміниться результативна ознака при зміні відповідного фактора на 1%. Середні показники еластичності можна порівнювати один з одним і відповідно ранжувати фактори за силою їхнього впливу на результативну ознаку.

### 10.5 Оцінка значущості множинної регресії і показники якості моделі

Практичну значущість рівняння множинної регресії оцінюють за допомогою **показника множинної кореляції** і його квадрата - **показника детермінації**.

Показник множинної кореляції характеризує тісноту зв'язку розглянутого набору факторів з досліджуваною ознакою або, інакше, оцінює тісноту загального впливу факторів на результативну ознаку.

Незалежно від форми зв'язку показник множинної кореляції може бути знайдений як індекс множинної кореляції:

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - \frac{\sigma_{\text{зал}}^2}{\sigma_y^2}}, \quad (10.36)$$

де  $\sigma_y^2$  – загальна дисперсія результативної ознаки;

$\sigma_{\text{зал}}^2$  – залишкова дисперсія.

Межі зміни індексу множинної кореляції від 0 до 1. Чим ближче його значення до 1, тим тісніше зв'язок результативної ознаки з усім набором досліджуваних факторів. Величина індексу множинної кореляції повинна бути більше (або дорівнювати) максимального значення парного індексу кореляції:

$$R_{yx_1x_2\dots x_m} \geq r_{yx_i(\max)}, \quad i = \overline{1, m}.$$

При правильному включенні факторів до регресійної моделі величина індексу множинної кореляції буде істотно відрізнятися від індексу кореляції парної залежності. Якщо ж додатково включені до рівняння множинної регресії фактори третьорядні, то індекс множинної кореляції може практично збігатися з індексом парної кореляції (розходження в третьому, четвертому знаках). Звідки випливає, що порівнюючи індекси множинної і парної кореляції, можна зробити висновок про доцільність включення до рівняння регресії того або іншого фактора.

Розрахунок індексу множинної кореляції передбачає визначення залишкової дисперсії:

$$\sigma_{\text{зал}}^2 = \frac{1}{n} \sum (y - \hat{y}_{x_1x_2\dots x_m})^2. \quad (10.37)$$

Можна користуватися формулою для індексу множинної детермінації:

$$R_{yx_1x_2\ldots x_m}^2 = 1 - \frac{\sum (y - \hat{y}_{x_1x_2\ldots x_m})^2}{\sum (y - \bar{y})^2}. \quad (10.38)$$

При лінійній залежності ознак формула для індексу множинної кореляції може бути представленою в такий спосіб:

$$R_{yx_1x_2\ldots x_m} = \sqrt{\sum_{i=1}^n \beta_i r_{yx_i}}, \quad (10.39)$$

де  $\beta_i$  – стандартизовані коефіцієнти регресії;

$r_{yx_i}$  – парні коефіцієнти кореляції результативної ознаки з кожним з факторів.

Вираз індексу множинної кореляції для лінійної регресії називають **лінійним коефіцієнтом множинної кореляції**, або сукупним коефіцієнтом кореляції.

Сукупний коефіцієнт кореляції можна також визначити за допомогою матриці парних коефіцієнтів кореляції:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}, \quad (10.40)$$

$$\text{де } \Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \cdots & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & \cdots & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & \cdots & r_{x_2x_p} \\ \cdots & \cdots & \cdots & 1 & \cdots \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & \cdots & 1 \end{vmatrix} \quad \text{- визначник матриці парних коефіцієнтів}$$

кореляції;

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_1} & \cdots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \cdots & r_{x_2x_p} \\ \cdots & \cdots & 1 & \cdots \\ r_{x_px_1} & r_{x_px_2} & \cdots & 1 \end{vmatrix} \quad \text{- визначник матриці міжфакторної кореляції.}$$

Як бачимо, величина множинного коефіцієнта кореляції залежить не тільки від кореляції результативної ознаки з кожним із факторів, але й від міжфакторної кореляції. Таким чином, визначити сукупний коефіцієнт кореляції можна, не звертаючись до рівняння множинної регресії, а використовуючи лише парні коефіцієнти кореляції.

У розглянутих показниках множинної кореляції (індекс і коефіцієнт) використовують значення залишкової дисперсії, що має систематичну помилку вбік зменшення. В результаті залишкова дисперсія тим менша, чим більше параметрів визначають у рівнянні регресії при заданому обсязі спостережень  $n$ . Якщо число параметрів при  $x_i$  дорівнює  $m$  і наближається до обсягу спостережень, то залишкова дисперсія буде близькою до нуля, і коефіцієнт (індекс) кореляції наблизиться до одиниці навіть при слабкому зв'язку факторів

з результативною ознакою. Для усунення цього явища використовують скорегований індекс (коефіцієнт) множинної кореляції.

**Скорегований індекс множинної кореляції** містить виправлення на число ступенів свободи, а саме залишкова сума квадратів  $\sum (y - \hat{y}_{x_1, x_2, \dots, x_m})^2$  ділиться на число ступенів свободи залишкової варіації  $n-m-1$ , а загальна сума квадратів відхилень  $\sum (y - \bar{y})^2$  - на число ступенів свободи в цілому за сукупністю  $(n-1)$ .

Нагадаємо, що  $R^2$  характеризує частку варіації залежної змінної, яка обумовлена регресією або мінливістю пояснюючих змінних. Чим ближче  $R^2$  до одиниці, тим краще регресія описує залежність між пояснюючими змінними і результативною ознакою. Разом з тим використання тільки одного коефіцієнта детермінації  $R^2$  для вибору найкращого рівняння регресії може виявитися недостатнім. На практиці зустрічаються випадки, коли погано визначена модель регресії може дати порівняно високий коефіцієнт  $R^2$ .

Недоліком коефіцієнта детермінації  $R^2$  також є те, що він збільшується при додаванні нових пояснюючих змінних у модель, хоча це і не обов'язково поліпшує якість регресійної моделі. Тому переважніше використовувати скорегований коефіцієнт детермінації  $\hat{R}^2$ , визначений за формулою

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1} (1 - R^2). \quad (10.41)$$

З (10.41) випливає, що чим більше кількість пояснюючих змінних  $p$ , тим менше  $\hat{R}^2$  порівняно з  $R^2$ . На відміну від  $R^2$  скорегований коефіцієнт  $\hat{R}^2$  може зменшуватися при введенні до моделі нових пояснюючих змінних, які не роблять істотного впливу на залежну змінну. Однак навіть збільшення скорегованого коефіцієнта детермінації  $\hat{R}^2$  при введенні до моделі нової пояснюючої змінної не завжди означає, що її коефіцієнт регресії є значущим (це відбувається тільки у випадку, якщо відповідне значення  $t$ -статистики більше за одиницю (за абсолютною величиною), тобто  $|t| > 1$ ). Інакше кажучи, збільшення  $\hat{R}^2$  ще не означає поліпшення якості регресійної моделі.

Як було показано вище, ранжирування факторів, що беруть участь у множинній лінійній регресії, може бути проведене через стандартизовані коефіцієнти регресії ( $\beta$ -коефіцієнти). Ця ж мета може бути досягнута за допомогою часткових коефіцієнтів кореляції (для лінійних зв'язків). Крім того, часткові показники кореляції широко використовують при розв'язанні проблеми відбору факторів: доцільність включення того або іншого фактора до моделі можна довести величиною показника часткової кореляції.

**Часткові коефіцієнти кореляції** характеризують тісноту зв'язку між результативною ознакою і відповідним фактором при елімінуванні (усуненні впливу) інших факторів, включених до рівняння регресії.

Показники часткової кореляції являють собою відношення зменшення залишкової дисперсії за рахунок додаткового включення до аналізу нового фактора до залишкової дисперсії, що мала місце до введення його в модель.

В загальному вигляді при наявності  $m$  факторів для рівняння

$$y = a + b_1x_1 + b_2x_2 + \dots + b_mx_m + \varepsilon$$

коефіцієнт часткової кореляції, що вимірює вплив на  $y$  фактору  $x_i$ , при незмінному рівні інших факторів, можна визначити за формулою:

$$r_{yx_1x_2\dots x_{i-1}x_{i+1}\dots x_m} = \sqrt{1 - \frac{1 - R_{yx_1x_2\dots x_i\dots x_m}^2}{1 - R_{yx_1x_2\dots x_{i-1}x_{i+1}\dots x_m}^2}}, \quad (10.42)$$

де  $R_{yx_1x_2\dots x_i\dots x_m}^2$  – множинний коефіцієнт детермінації всіх  $m$  факторів з результативною ознакою;

$R_{yx_1x_2\dots x_{i-1}x_{i+1}\dots x_m}^2$  – той самий показник детермінації, але без введення до моделі фактору  $x_i$ .

При двох факторах формула (10.42) прийме вигляд:

$$r_{yx_1x_2} = \sqrt{1 - \frac{1 - R_{yx_1x_2}^2}{1 - R_{yx_2}^2}}, \quad r_{yx_2x_1} = \sqrt{1 - \frac{1 - R_{yx_1x_2}^2}{1 - R_{yx_1}^2}}. \quad (10.43)$$

Порядок часткового коефіцієнта кореляції визначають кількістю факторів, вплив яких виключається. Наприклад,  $r_{yx_1x_2}$  – коефіцієнт часткової кореляції першого порядку. Відповідно коефіцієнти парної кореляції називають коефіцієнтами нульового порядку. Коефіцієнти часткової кореляції більш високих порядків можна визначити через коефіцієнти часткової кореляції більш низьких порядків за рекурентною формулою:

$$r_{yx_ix_1x_2\dots x_{i-1}x_{i+1}\dots x_m} = \frac{r_{yx_ix_1x_2\dots x_{i-1}x_{i+1}\dots x_{m-1}} - r_{yx_mx_1x_2\dots x_{m-1}} * r_{yx_ix_mx_1x_2\dots x_{i-1}x_{i+1}\dots x_{m-1}}}{\sqrt{(1 - r_{yx_mx_1x_2\dots x_{m-1}}^2) * (1 - r_{yx_ix_mx_1x_2\dots x_{i-1}x_{i+1}\dots x_{m-1}}^2)}}. \quad (10.44)$$

Для двох факторів дана формула прийме вигляд:

$$r_{yx_1x_2} = \frac{r_{yx_1} - r_{yx_2} * r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2) * (1 - r_{x_1x_2}^2)}}, \quad r_{yx_2x_1} = \frac{r_{yx_2} - r_{yx_1} * r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2) * (1 - r_{x_1x_2}^2)}}. \quad (10.45)$$

Для рівняння регресії з трьома факторами часткові коефіцієнти кореляції другого порядку визначають на основі часткових коефіцієнтів кореляції першого порядку. Так, за рівнянням  $y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon$  можливе вираховування трьох часткових коефіцієнтів кореляції другого порядку:

$$r_{yx_1x_2x_3}, r_{yx_2x_1x_3}, r_{yx_3x_1x_2},$$

кожний з яких визначають за рекурентною формулою. Наприклад, при  $i=1$  маємо формулу для розрахунку  $r_{yx_1x_2x_3}$ :

$$r_{yx_1x_2x_3} = \frac{r_{yx_1x_2} - r_{yx_3x_2} * r_{x_1x_2x_3}}{\sqrt{(1 - r_{yx_3x_2}^2) * (1 - r_{x_1x_2x_3}^2)}}. \quad (10.46)$$

Розраховані за рекурентною формулою часткові коефіцієнти кореляції змінюються в межах від  $-1$  до  $+1$ , а за формулами через множинні коефіцієнти детермінації – від  $0$  до  $1$ . Порівняння їх одного з одним дозволяє ранжувати

фактори за тіснотою їхнього зв'язку з результативною ознакою. Часткові коефіцієнти кореляції дають міру тісноти зв'язку кожного фактора з результативною ознакою в чистому вигляді. Якщо із стандартизованого рівняння регресії  $t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \beta_3 t_{x_3} + \varepsilon$  випливає, що  $\beta_1 > \beta_2 > \beta_3$ , тобто за силою впливу на результативну ознаку порядок факторів такий:  $x_1, x_2, x_3$ , то цей же порядок факторів визначають і за співвідношенням часткових коефіцієнтів кореляції,  $r_{yx_1x_2x_3} > r_{yx_2x_1x_3} > r_{yx_3x_1x_2}$ .

В економетриці часткові коефіцієнти кореляції звичайно не мають самостійного значення. Їх використовують на стадії формування моделі. Так, будуючи багатофакторну модель, на першому кроці визначають рівняння регресії з повним набором факторів і розраховують матрицю часткових коефіцієнтів кореляції. На другому кроці відбирають фактор з найменшою і несуттєвою за  $t$ -критерієм Стюдента величиною показника часткової кореляції. Виключивши його з моделі, будують нове рівняння регресії. Процедура триває доти, поки не виявиться, що всі часткові коефіцієнти кореляції істотно відрізняються від нуля. Якщо виключено несуттєвий фактор, то множинні коефіцієнти детермінації на двох суміжних кроках побудови регресійної моделі майже не відрізняються один від одного,  $R_{m+1}^2 \approx R_m^2$ , де  $m$  – число факторів.

З наведених вище формул часткових коефіцієнтів кореляції видний зв'язок цих показників із сукупним коефіцієнтом кореляції. Знаючи часткові коефіцієнти кореляції (послідовно першого, другого і більш високого порядку), можна визначити сукупний коефіцієнт кореляції за формулою:

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - (1 - r_{yx_1}^2) * (1 - r_{yx_2x_1}^2) * (1 - r_{yx_3x_1x_2}^2) * \dots * (1 - r_{yx_mx_1x_2\dots x_{m-1}}^2)}. \quad (10.47)$$

Зокрема, для двофакторного рівняння формула (10.47) приймає вигляд:

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - (1 - r_{yx_1}^2) * (1 - r_{yx_2x_1}^2)}. \quad (10.48)$$

При повній залежності результативної ознаки від досліджуваних факторів коефіцієнт їхнього сукупного впливу дорівнює одиниці. Від одиниці віднімають частку залишкової варіації результативної ознаки  $(1 - r^2)$ , що зумовлена послідовно включеними до аналізу факторами. В результаті підкореневий вираз характеризує сукупну дію всіх досліджуваних факторів.

Значущість рівняння множинної регресії в цілому, так само як і в парній регресії, оцінюють за допомогою  $F$ -критерію Фішера:

$$F = \frac{S_{\text{факт}}}{S_{\text{зал}}} = \frac{R^2}{1 - R^2} * \frac{n - m - 1}{m}, \quad (10.49)$$

де  $S_{\text{факт}}$  – факторна сума квадратів на один ступінь свободи;

$S_{\text{зал}}$  – залишкова сума квадратів на один ступінь свободи;

$R^2$  – коефіцієнт (індекс) множинної детермінації;

$m$  – число параметрів при змінних  $x$  (в лінійній регресії збігається з числом включених до моделі факторів);

$n$  – число спостережень.

Оцінюють значущість не тільки рівняння в цілому, але й фактора, додатково включеного до регресійної моделі. Необхідність такої оцінки пов'язана з тим, що не кожен фактор, що ввійшов до моделі, може істотно збільшувати частку поясненої варіації результативної ознаки. Крім того, за наявності в моделі декількох факторів, їх можна вводити до моделі в різній послідовності. Через кореляцію між факторами значущість того самого фактора може бути різною залежно від послідовності його введення до моделі. Мірою для оцінки включення фактора до моделі служить частковий  $F$ -критерій, тобто  $F_{x_i}$ .

Частковий  $F$ -критерій побудований на порівнянні приросту факторної дисперсії, обумовленого впливом додатково включеного фактора, із залишковою дисперсією на один ступінь свободи за регресійною моделлю в цілому. В загальному вигляді для фактора  $x_i$  частковий  $F$ -критерій визначиться як

$$F_{x_i} = \frac{R^2_{yx_1 \dots x_i \dots x_m} - R^2_{yx_1 \dots x_{i-1} x_{i+1} \dots x_m}}{1 - R^2_{yx_1 \dots x_i \dots x_m}} * \frac{n - m - 1}{1}, \quad (10.50)$$

де  $R^2_{yx_1 \dots x_i \dots x_m}$  – коефіцієнт множинної детермінації для моделі з повним набором факторів;

$R^2_{yx_1 \dots x_{i-1} x_{i+1} \dots x_m}$  – той же показник, але без включення до моделі фактора  $x_i$ ;

$n$  – число спостережень;

$m$  – число параметрів у моделі (без вільного члена).

Фактичне значення часткового  $F$ -критерію порівнюють з табличним за рівнем значущості  $\alpha$  і числом ступенів свободи: 1 і  $n-m-1$ . Якщо фактичне значення  $F_{x_i}$  перевищує  $F(\alpha, k_1, k_2)$ , то додаткове включення фактора  $x_i$  до моделі статистично виправдане і коефіцієнт чистої регресії  $b_i$  при факторі  $x_i$  статистично значущий. Якщо ж фактичне значення  $F_{x_i}$  менше табличного, то додаткове включення до моделі фактора  $x_i$  не збільшує істотно частку поясненої варіації ознаки  $y$ , отже, недоцільно його включення до моделі; коефіцієнт регресії при даному факторі в цьому випадку статистично не значущий.

Для двохфакторного рівняння часткові  $F$ -критерії мають вигляд:

$$F_{x_1} = \frac{R^2_{yx_1 x_2} - r^2_{yx_2}}{1 - R^2_{yx_1 x_2}} * (n - 3), \quad F_{x_2} = \frac{R^2_{yx_1 x_2} - r^2_{yx_1}}{1 - R^2_{yx_1 x_2}} * (n - 3). \quad (10.51)$$

За допомогою часткового  $F$ -критерію можна перевірити значущість всіх коефіцієнтів регресії в припущенні, що кожний відповідний фактор  $x_i$  вводився до рівняння множинної регресії останнім.

Частковий  $F$ -критерій оцінює значущість коефіцієнтів чистої регресії. Знаючи величину  $F_{x_i}$ , можна визначити і  $t$ -критерій для коефіцієнта регресії при  $i$ -му факторі,  $t_{b_i}$ , а саме:

$$t_{b_i} = \sqrt{F_{x_i}}. \quad (10.52)$$



Оцінка значущості коефіцієнтів чистої регресії за  $t$ -критерієм Стюдента може бути проведена і без розрахунку часткових  $F$ -критеріїв. В цьому випадку, як і в парній регресії, для кожного фактора використовують формулу:

$$t_{b_i} = \frac{b_i}{m_{b_i}}, \quad (10.53)$$

де  $b_i$  – коефіцієнт чистої регресії при факторі  $x_i$ ;

$m_{b_i}$  – середня квадратична (стандартна) помилка коефіцієнта регресії  $b_i$ .

Для рівняння множинної регресії  $\hat{y} = a + b_1x_1 + b_2x_2 + \dots + b_mx_m$  середню квадратичну помилку коефіцієнта регресії можна визначити за наступною формулою:

$$m_{b_i} = \frac{\sigma_y \sqrt{1 - R_{yx_1 \dots x_m}^2}}{\sigma_{x_i} \sqrt{1 - R_{x_i x_1 \dots x_m}^2}} \frac{1}{\sqrt{n - m - 1}}, \quad (10.54)$$

де  $\sigma_y$  – середнє квадратичне відхилення для ознаки  $y$ ;

$\sigma_{x_i}$  – середнє квадратичне відхилення для ознаки  $x_i$ ;

$R_{yx_1 \dots x_m}^2$  – коефіцієнт детермінації для рівняння множинної регресії;

$R_{x_i x_1 \dots x_m}^2$  – коефіцієнт детермінації для залежності фактору  $x_i$  з усіма іншими факторами рівняння множинної регресії;

$n - m - 1$  – число ступенів свободи для залишкової суми квадратів відхилень.

Як бачимо, щоб скористатися даною формулою, необхідно використати матрицю міжфакторної кореляції і розрахувати за нею відповідні коефіцієнти детермінації  $R_{x_i x_1 \dots x_m}^2$ . Так, для рівняння  $\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3$  оцінка значущості коефіцієнтів регресії  $b_1$ ,  $b_2$ ,  $b_3$  припускає розрахунок трьох міжфакторних коефіцієнтів детермінації:  $R_{x_1 x_2 x_3}^2$ ,  $R_{x_2 x_1 x_3}^2$ ,  $R_{x_3 x_1 x_2}^2$ .

Взаємозв'язок показників часткового коефіцієнта кореляції, часткового  $F$ -критерію і  $t$ -критерію Стюдента для коефіцієнтів чистої регресії може використовуватися в процедурі відбору факторів. Відсівання факторів при побудові рівняння регресії методом виключення практично можна здійснювати не тільки за частковими коефіцієнтами кореляції, виключаючи на кожному кроці фактор з найменшим незначущим значенням часткового коефіцієнта кореляції, але й за величинами  $t_{b_i}$  і  $F_{x_i}$ . Частковий  $F$ -критерій широко використовують і при побудові моделі за методом включення змінних і за кроковим регресійним методом.

## Контрольні запитання

1. Поясніть терміни «рівняння регресії» і «функція регресії».
2. Які помилки належать до помилок специфікації? До помилок виміру?
3. На яких підставах вибирають вид регресійної залежності?
4. Як розраховують залишкову дисперсію? Загальну дисперсію ознаки  $y$ ?

5. У чому полягає сутність методу найменших квадратів?
6. Якими властивостями повинна володіти лінійна модель, щоб оцінки її параметрів мали найменшу дисперсію в класі всіх лінійних незміщених оцінок?
7. Поясніть, що таке коефіцієнт регресії? Коефіцієнт кореляції? Коефіцієнт детермінації?
8. Як визначають статистичну значущість рівняння регресії в цілому?
9. З якою метою визначають стандартну помилку коефіцієнта регресії? Який критерій для цього використовують?
10. Як визначають довірчий інтервал для коефіцієнта регресії?
11. З якою метою і як визначають значущість лінійного коефіцієнта кореляції?
12. У якому випадку використовують лінійну модель множинної регресії?
13. Що таке коефіцієнти «чистої» регресії? Який в них економічний зміст?
14. У якому випадку використовують стандартизовані коефіцієнти регресії?
15. Що характеризують показник множинної кореляції і показник детермінації?
16. Що таке лінійний коефіцієнт множинної кореляції і скорегований індекс множинної кореляції?
17. Для чого розраховують часткові коефіцієнти кореляції?
18. Як оцінюють значущість рівняння множинної регресії в цілому та значущість коефіцієнтів чистої регресії?
19. Поясніть, у чому полягає процедура відбору факторів при побудові рівняння регресії методом виключення?

## **ТЕМА 11**

### **МУЛЬТИКОЛІНЕАРНІСТЬ І ЇЇ ВПЛИВ НА ОЦІНКИ ПАРАМЕТРІВ МОДЕЛІ**

#### **11.1 Поняття мультиколінеарності.**

##### **Вплив мультиколінеарності на оцінки параметрів**

Однією з проблем, пов'язаних з практичним використанням моделі множинної регресії, є мультиколінеарність. Під мультиколінеарністю розуміють високу взаємну корельованість пояснюючих змінних. Мультиколінеарність може проявлятися у функціональній (явній) і стохастичній (схованій) формах.

При функціональній формі мультиколінеарності принаймні один з парних зв'язків між пояснюючими змінними є лінійною функціональною залежністю. Це призводить до неможливості розв'язання відповідної системи нормальних рівнянь і одержання оцінок параметрів регресійної моделі.

Однак в економічних дослідженнях мультиколінеарність частіше проявляється в стохастичній формі, коли між хоча б двома пояснюючими змінними існує тісний кореляційний зв'язок. Оцінки параметрів стають дуже

чутливими до незначної зміни результатів спостережень і обсягу вибірки. Рівняння регресії в цьому випадку, як правило, не мають реального змісту, тому що деякі з коефіцієнтів можуть мати неправильні з погляду економічної теорії знаки і невиправдано великі значення.

Точних кількісних критеріїв для визначення наявності або відсутності мультиколінеарності не існує. Проте, є деякі евристичні підходи до її виявлення. Один з таких підходів полягає в аналізі кореляційної матриці між пояснюючими змінними  $x_j$  і виявленні пар змінних, що мають високі коефіцієнти кореляції (звичайно більші за 0,8). Якщо такі змінні існують, то говорять про мультиколінеарність між ними.

Корисно також знаходити множинні коефіцієнти детермінації між однією з пояснюючих змінних і деякою групою з них. Наявність високого множинного коефіцієнта детермінації (звичайно більшого за 0,6) свідчить про мультиколінеарність.

Для усунення або зменшення мультиколінеарності використовують ряд методів. Найпростіший з них (але далеко не завжди можливий) полягає в тому, що із двох пояснюючих змінних, які мають високий коефіцієнт кореляції (більший за 0,8), одну змінну виключають з розгляду. При цьому яку змінну залишити, а яку видалити з аналізу, вирішують у першу чергу на підставі економічних міркувань. Якщо з економічної точки зору ні одній зі змінних не можна надати перевагу, то залишають ту із двох змінних, котра має більший коефіцієнт кореляції із залежною змінною. У цій вимозі проявляється специфіка множинної регресії як методу дослідження комплексного впливу факторів в умовах їхньої незалежності один від одного.

За величиною парних коефіцієнтів кореляції виявляється лише явна колінеарність факторів. Найбільші труднощі у використанні апарата множинної регресії виникають при наявності мультиколінеарності факторів, коли більш ніж два фактори зв'язані між собою лінійною залежністю, тобто має місце сукупний вплив факторів один на одного. Наявність мультиколінеарності факторів може означати, що деякі фактори будуть завжди діяти в унісон. У результаті варіація у вихідних даних перестає бути повністю незалежною і не можна оцінити вплив кожного фактора окремо.

Включення до моделі мультиколінеарних факторів небажане в силу наступних наслідків:

- утруднюється інтерпретація параметрів множинної регресії як характеристик дії факторів в «чистому» вигляді, тому що фактори корельовані; параметри лінійної регресії втрачають економічний зміст;
- оцінки параметрів ненадійні, виявляють великі стандартні помилки й змінюються зі зміною обсягу спостережень (не тільки за величиною, але й за знаком), що робить модель непридатною для аналізу й прогнозування.

Для оцінки мультиколінеарності факторів може використовуватися визначник матриці парних коефіцієнтів кореляції між факторами.

Якщо фактори не корельовані один з одним, то матриця парних коефіцієнтів кореляції між факторами є одиничною матрицею, оскільки всі

недіагональні елементи  $r_{x_i y_j}$  ( $i \neq j$ ) дорівнюють нулю. Так, для рівняння, що включає три пояснюючих змінних

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + b_3 x_m$$

матриця коефіцієнтів кореляції між факторами має визначник, що дорівнює одиниці:

$$\text{Det} = \begin{vmatrix} r_{x_1 x_1} & r_{x_1 x_2} & r_{x_1 x_3} \\ r_{x_2 x_1} & r_{x_2 x_2} & r_{x_2 x_3} \\ r_{x_3 x_1} & r_{x_3 x_2} & r_{x_3 x_3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1.$$

Якщо ж між факторами існує повна лінійна залежність, і всі коефіцієнти кореляції дорівнюють одиниці, то визначник такої матриці дорівнює нулю:

$$\text{Det} = \begin{vmatrix} r_{x_1 x_1} & r_{x_1 x_2} & r_{x_1 x_3} \\ r_{x_2 x_1} & r_{x_2 x_2} & r_{x_2 x_3} \\ r_{x_3 x_1} & r_{x_3 x_2} & r_{x_3 x_3} \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

Чим ближче до нуля визначник матриці міжфакторної кореляції, тим сильніша мультиколінеарність факторів і ненадійніші результати множинної регресії. І, навпаки, чим ближче до одиниці визначник матриці міжфакторної кореляції, тим менше мультиколінеарність факторів.

## 11.2 Методи виключення мультиколінеарності

Існує ряд підходів до подолання сильної міжфакторної кореляції. Найпростіший шлях усунення мультиколінеарності полягає у виключенні з моделі одного або кількох факторів. Інший підхід пов'язаний з перетворенням факторів, за яким зменшується кореляція між ними.

Одним із шляхів урахування внутрішньої кореляції факторів є перехід до сполучених рівнянь регресії, тобто до рівнянь, які відбивають не тільки вплив факторів, але і їхню взаємодію. Так, якщо  $\hat{y} = a + b_1 x_1 + b_2 x_2 + b_3 x_m$ , то можлива побудова сполученого рівняння наступного вигляду:

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + \varepsilon.$$

Розглянуте рівняння включає взаємодію першого порядку (взаємодія двох факторів). Можливе включення до моделі і взаємодій більш високого порядку, якщо буде доведена їхня статистична значущість за  $F$ -критерієм Фішера, але, як правило, взаємодії третього й більш високих порядків виявляються статистично незначущими.

Інший метод усунення або зменшення мультиколінеарності полягає в переході від незміщених оцінок, визначених за методом найменших квадратів, до зміщених оцінок, що володіють, однак, меншим розсіюванням оцінюваного параметра. Оцінки параметрів множинної регресії володіють відповідно до теореми Гауса-Маркова мінімальними дисперсіями в класі всіх лінійних незміщених оцінок, але при наявності мультиколінеарності ці дисперсії можуть виявитися занадто більшими, і звертання до відповідних зміщених оцінок може підвищити точність оцінювання параметрів регресії.

Для усунення мультиколінеарності може бути використаний перехід від вихідних пояснюючих змінних  $x_1, x_2, \dots, x_n$ , зв'язаних між собою досить тісною кореляційною залежністю, до нових змінних, які є лінійними комбінаціями вихідних. При цьому нові змінні повинні бути слабко корельованими або взагалі некорельованими. У якості таких змінних беруть, наприклад, так звані головні компоненти вектора вихідних пояснюючих змінних, досліджувані в компонентному аналізі, і розглядають регресію на головних компонентах, у якій останні виступають у якості узагальнених пояснюючих змінних, що підлягають надалі змістовній (економічній) інтерпретації.

Ортогональність головних компонентів запобігає прояві ефекту мультиколінеарності. Крім того, застосовуваний метод дозволяє обмежитися малим числом головних компонентів при порівняно великій кількості вихідних пояснюючих змінних.

Ще одним з можливих методів усунення або зменшення мультиколінеарності є використання покрокових процедур відбору найбільш інформативних змінних. Наприклад, на першому кроці розглядається лише одна пояснююча змінна, що має із залежною змінною у найбільший коефіцієнт детермінації. На другому кроці включається до регресії нова пояснююча змінна, котра разом з першою утворює пару пояснюючих змінних, що має з у найбільш високий (скорегований) коефіцієнт детермінації. На третьому кроці вводиться до регресії ще одна пояснююча змінна, котра разом із двома відібраними утворює трійку пояснюючих змінних, що має з у найбільший (скорегований) коефіцієнт детермінації, та ін.

Процедура введення нових змінних триває доти, поки буде збільшуватися відповідний (скорегований) коефіцієнт детермінації  $R^2$  (точніше - мінімальне значення  $R^2_{\min}$ ).

Розглянемо приклад. За даними  $n=20$  сільськогосподарських районів області досліджується залежність змінної  $y$  - урожайності зернових культур (у ц/га) від ряду змінних - факторів сільськогосподарського виробництва:  $x_1$  - число тракторів (приведеної потужності на 100 га);  $x_2$  - число зернозбиральних комбайнів на 100 га;  $x_3$  - число знарядь поверхневої обробки ґрунту на 100 га;  $x_4$  - кількість добрив, що витрачаються на 1 га (т/га);  $x_5$  - кількість хімічних засобів захисту рослин, що витрачаються на 1 га (ц/га). Вихідні дані наведені в таблиці 11.1.

Таблиця 11.1 – Вихідні дані до прикладу

<b>i (номер району)</b>	<b><math>y_i</math></b>	<b><math>x_{i1}</math></b>	<b><math>x_{i2}</math></b>	<b><math>x_{i3}</math></b>	<b><math>x_{i4}</math></b>	<b><math>x_{i5}</math></b>
1	9,70	1,59	0,26	2,05	0,32	0,14
2	8,40	0,34	0,28	0,46	0,59	0,66
19	13,10	0,08	0,25	0,03	0,73	0,20
20	8,70	1,36	0,26	0,17	0,99	0,42

У випадку виявлення мультиколінеарності вжити заходи з її усунення (зменшення), використовуючи покрокову процедуру відбору найбільш інформативних змінних.

З системи (10.26) знайдемо вектор оцінок параметрів регресійної моделі  $b = (3,515; -0,006; 15,542; 0,110; 4,475; -2,932)$ , так що відповідно до (10.24) вибіркове рівняння множинної регресії має вигляд:

$$y = 33,515 - 0,006x_1 + 15,542x_2 + 0,110x_3 + 4,475x_4 - 2,932x_5,$$

де середні квадратичні відхилення (стандартні помилки) коефіцієнтів регресії  $b_j$  обчислені за формулою (10.54) відповідно дорівнюють 5,41; 0,60; 21,59; 0,85; 1,54; 3,09.

Порівнюючи значення  $t$ -статистики (за абсолютною величиною) кожного коефіцієнта регресії  $b_j$  за формулою

$$t_{bj} = \frac{b_j}{\sigma_{bj}}, j = (0, 1, 2, 3, 4, 5),$$

тобто  $t_{b0} = 0,65$ ,  $t_{b1} = -0,01$ ,  $t_{b2} = 0,72$ ,  $t_{b3} = 0,13$ ,  $t_{b4} = 2,91$ ,  $t_{b5} = -0,95$  із критичним значенням  $t_{0,95;14} = 2,14$ , визначеним з довідкової таблиці на рівні значущості  $\alpha = 0,05$  при числі ступенів свободи  $k = n - p - 1 = 20 - 5 - 1 = 14$ , бачимо, що значущим виявився тільки коефіцієнт регресії  $b_4$  при змінній  $x_4$  - кількість добрив, що витрачаються на гектар землі.

Множинний коефіцієнт детермінації врожайності зернових культур  $y$  за сукупністю п'яти факторів ( $x_1$ - $x_5$ ) сільськогосподарського виробництва виявився рівним  $R^2_{y12345} = 0,517$ , тобто 51,7% варіації залежної змінної пояснюється включеними до моделі п'ятьма пояснюючими змінними. Оскільки обчислене фактичне значення  $F = 3,00$  перевищує табличне  $F_{0,05;5;14} = 2,96$ , то рівняння регресії значуще за  $F$ -критерієм на рівні  $\alpha = 0,05$ .

Розраховано матрицу парних коефіцієнтів кореляції, що наведена у таблиці 11.2.

Таблиця 11.2 - Матриця парних коефіцієнтів кореляції

Змінні	$y$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
$y$	1,00	0,43	0,37	0,40	0,58*	0,33
$x_1$	0,43	1,00	0,85*	0,98*	0,11	0,34
$x_2$	0,37	0,85*	1,00	0,88*	0,03	0,46*
$x_3$	0,40	0,98*	0,88*	1,00	0,03	0,28
$x_4$	0,58*	0,11	0,03	0,03	1,00	0,57*
$x_5$	0,33	0,34	0,46*	0,28	0,57*	1,00

Знаком \* відзначені коефіцієнти кореляції, значущі за  $t$ -критерієм на 5%-му рівні.

Аналізуючи матрицу парних коефіцієнтів кореляції, можна відзначити тісний кореляційний зв'язок між змінними  $x_1$  і  $x_2$  ( $r_{12} = 0,85$ ),  $x_1$  і  $x_3$  ( $r_{13} = 0,98$ ),  $x_2$  і  $x_3$  ( $r_{23} = 0,88$ ), що, свідчить про мультиколінеарність пояснюючих змінних.

Для усунення мультиколінеарності застосуємо процедуру покрокового відбору найбільш інформативних змінних.

**1-й крок.** З пояснюючих змінних  $x_1$ - $x_5$  виділяємо змінну  $x_4$ , що має із залежною змінною  $y$  найбільший коефіцієнт детермінації  $R^2_{y4}$  (рівний для парної моделі квадрату коефіцієнта кореляції  $r^2_{y4}$ ). Очевидно, це змінна  $x_4$ , тому що коефіцієнт детермінації  $R^2_{y4} = r^2_{y4} = 0,58^2 = 0,336$  — максимальний. З урахуванням виправлення на незміщеність за формулою скорегований коефіцієнт детермінації  $\hat{R}^2_{y4} = 1 - \frac{19}{18}(1 - 0,336) = 0,299$ .

**2-й крок.** Серед усіляких пар пояснюючих змінних  $x_4, x_j, j = 1,2,3,5$ , обирається пара  $(x_4, x_3)$ , що має із залежною змінною  $y$  найбільш високий коефіцієнт детермінації  $R^2_{y43} = R^2_{y34} = 0,483$  й з урахуванням виправлення  $\hat{R}^2_{y43} = 1 - \frac{19}{17}(1 - 0,483) = 0,422$ .

**3-й крок.** Серед усіляких трійок пояснюючих змінних  $(x_4, x_3, x_j), j = 1,2,5$ , найбільш інформативною виявилася трійка  $(x_4, x_3, x_5)$ , що має максимальний коефіцієнт детермінації  $R^2_{y435} = R^2_{y543} = 0,513$  й відповідно скорегований коефіцієнт  $\hat{R}^2_{y435} = 0,422$ .

Оскільки скорегований коефіцієнт детермінації на 3-му кроці не збільшився, то в регресійній моделі досить обмежитися лише двома відібраними раніше пояснюючими змінними  $x_4$  і  $x_3$ .

Рівняння регресії по цим змінним прийме вигляд:

$$\hat{y} = 7,29 + 3,48x_3 + 3,48x_4,$$

де середні квадратичні відхилення коефіцієнтів регресії  $b_i$  обчислені за формулою (10.54) відповідно дорівнюють 0,66; 0,13; 1,07.

Неважко переконатися в тому, що тепер всі коефіцієнти регресії значущі, тому що кожне зі значень  $t$ -статистики

$$t_{b0} = \frac{7,29}{0,66} = 11,0; \quad t_{b3} = \frac{3,48}{0,13} = 26,8; \quad t_{b4} = \frac{3,48}{1,07} = 3,25$$

більше відповідного табличного значення  $t_{0,95;17} = 2,11$ .

Оскільки значення коефіцієнтів кореляції досить високі (понад 0,8):  $r_{12} = 0,85, r_{13} = 0,98, r_{23} = 0,88$ , очевидно, що з відповідних трьох змінних  $x_1, x_2, x_3$  дві змінні можна було відразу виключити з регресії й без проведення покрокового відбору, але як саме змінні виключити — треба вирішувати, виходячи з якісних міркувань, заснованих на знанні предметної області (у цьому випадку впливу на врожайність факторів сільськогосподарського виробництва).

Крім розглянутої вище покрокової процедури приєднання пояснюючих змінних використовуються також покрокові процедури приєднання-видалення і процедура видалення пояснюючих змінних. Слід зазначити, що яка би покрокова процедура не використовувалася, вона не гарантує визначення оптимального (у смислі одержання максимального коефіцієнта детермінації  $R^2$ ) набору пояснюючих змінних. Однак у більшості випадків одержувані за допомогою покрокових процедур набори змінних виявляються оптимальними або близькими до оптимальних.

Відбір факторів, що включаються до регресії, є одним з найважливіших етапів практичного використання методів регресії. Підходи до відбору факторів на основі показників кореляції можуть бути різними. Вони приводять побудову рівняння множинної регресії відповідно до різних методик. Залежно від того, яка методика побудови рівняння регресії прийнята, змінюється алгоритм її рішення на ЕОМ.

При відборі факторів рекомендують також користуватися наступним правилом: число факторів, що включають до моделі, звичайно в 6-7 разів менше обсягу сукупності, за якою будують регресію. Якщо це співвідношення порушено, то число ступенів свободи залишкової дисперсії дуже мале. Це призводить до того, що параметри рівняння регресії виявляються статистично незначущими, а  $F$ -критерій менше за табличне значення.

### Контрольні запитання

1. Поясніть поняття мультиколінеарності та у яких формах вона може проявлятися?
2. Які критерії використовують для визначення наявності або відсутності мультиколінеарності?
3. Охарактеризуйте методи, використовувані для усунення або зменшення мультиколінеарності.
4. Поясніть, у чому полягають наслідки включення до моделі мультиколінеарних факторів.
5. Як для оцінки мультиколінеарності факторів використовують визначник матриці парних коефіцієнтів кореляції між факторами?
6. Поясніть алгоритм покрокових процедур відбору найбільш інформативних змінних.

## ТЕМА 12

### УЗАГАЛЬНЕНИЙ МЕТОД НАЙМЕНШИХ КВАДРАТІВ

#### 12.1 Поняття гомоскедастичності і гетероскедастичності

При оцінці параметрів рівняння регресії застосовують метод найменших квадратів (МНК). При цьому відповідно до теореми Гауса-Маркова виконують певні передумови щодо випадкової складової  $\varepsilon$ . У моделі

$$y = a + b_1x_1 + b_2x_2 + \dots + b_mx_m + \varepsilon$$

випадкова складова  $\varepsilon$  є неспостережуваною величиною. Після того як зроблено оцінку параметрів моделі, розраховуючи різниці фактичних і теоретичних значень результативної ознаки  $y$ , можна визначити оцінки випадкової складової  $y - \hat{y}_x$ . Їх можна вважати деякою вибірковою реалізацією невідомого залишку заданого рівняння, тобто  $\varepsilon_i$ .



При зміні специфікації моделі, додаванні до неї нових спостережень вибіркові оцінки залишків  $\varepsilon_i$  можуть змінюватися. Тому до задачі регресійного аналізу входить не тільки побудова самої моделі, але й дослідження випадкових відхилень  $\varepsilon_i$ , тобто залишкових величин.

При використанні критеріїв Фішера і Стюдента роблять припущення щодо поведінки залишків  $\varepsilon_i$ , тобто що вони являють собою незалежні випадкові величини і їхнє математичне сподівання дорівнює 0, що вони мають однакову (постійну) дисперсію і підкоряються нормальному розподілу.

Статистична перевірка параметрів регресії і показників кореляції заснована на неперевіряємих передумовах розподілу випадкової складової  $\varepsilon_i$ . Вона носить лише попередній характер. Після побудови рівняння регресії проводять перевірку наявності у випадкових залишків  $\varepsilon_i$  передбачуваних властивостей. Це пов'язане з тим, що оцінки параметрів регресії повинні бути незміщеними, спроможними і ефективними. **Незміщеність** оцінки означає, що математичне сподівання залишків дорівнює нулю. Якщо оцінки параметрів мають властивість незміщеності, то їх можна порівнювати. Оцінки вважаються **ефективними**, якщо вони характеризуються найменшою дисперсією. У практичних дослідженнях це означає можливість переходу від точкового оцінювання параметрів до інтервального. **Спроможність** оцінок характеризує збільшення їхньої точності зі збільшенням обсягу вибірки. Великий практичний інтерес представляють ті результати регресійного аналізу, для яких довірчий інтервал очікуваного значення параметра регресії  $b_i$  має у границі значення імовірності, що дорівнює одиниці. Іншими словами, імовірність одержання оцінки на заданій відстані від істинного значення параметра близька до одиниці.

Метод найменших квадратів визначає оцінки регресії на основі мінімізації суми квадратів залишків. Тому дослідження залишків  $\varepsilon_i$  припускає перевірку наявності наступних п'яти передумов МНК:

- збурення  $\varepsilon$  є величина випадкова;
- математичне сподівання збурення  $\varepsilon$  дорівнює нулю;
- дисперсія збурення  $\varepsilon$  постійна для будь-якого  $i$  - умова гомоскедастичності або рівномірності збурення;
- відсутність автокореляції залишків – значення залишків  $\varepsilon_i$  розподілені незалежно один від одного;
- збурення  $\varepsilon_i$  є нормально розподіленою випадковою величиною.

Якщо розподіл випадкових залишків  $\varepsilon_i$  не відповідає деяким передумовам МНК, то необхідно корегувати модель. Насамперед, перевіряють випадковий характер залишків  $\varepsilon_i$  – перша передумова МНК. З цією метою будують графік залежності залишків  $\varepsilon_i$  від теоретичних значень результативної ознаки (рис. 12.1, а). Якщо графіком є горизонтальна смуга точок, то залишки  $\varepsilon_i$

являють собою випадкові величини і теоретичні значення  $\hat{y}_x$  добре апроксимують фактичні значення  $y$ .

Друга передумова МНК щодо нульового значення математичного сподівання залишків означає, що  $\sum (y - \hat{y}_x) = 0$ . Це здійснимо для лінійних моделей і моделей, що нелінійні відносно включених змінних.

Разом з тим, незміщеність оцінок коефіцієнтів регресії, отриманих за МНК, залежить від незалежності випадкових залишків і величин  $x_j$ , що також досліджують в рамках дотримання другої передумови МНК. З цією метою поряд з розглянутим графіком залежності залишків  $\varepsilon_i$  від теоретичних значень результативної ознаки  $\hat{y}_x$  будують графік залежності випадкових залишків  $\varepsilon_i$  від факторів, включених до регресії  $x_j$  (рис. 12.2).

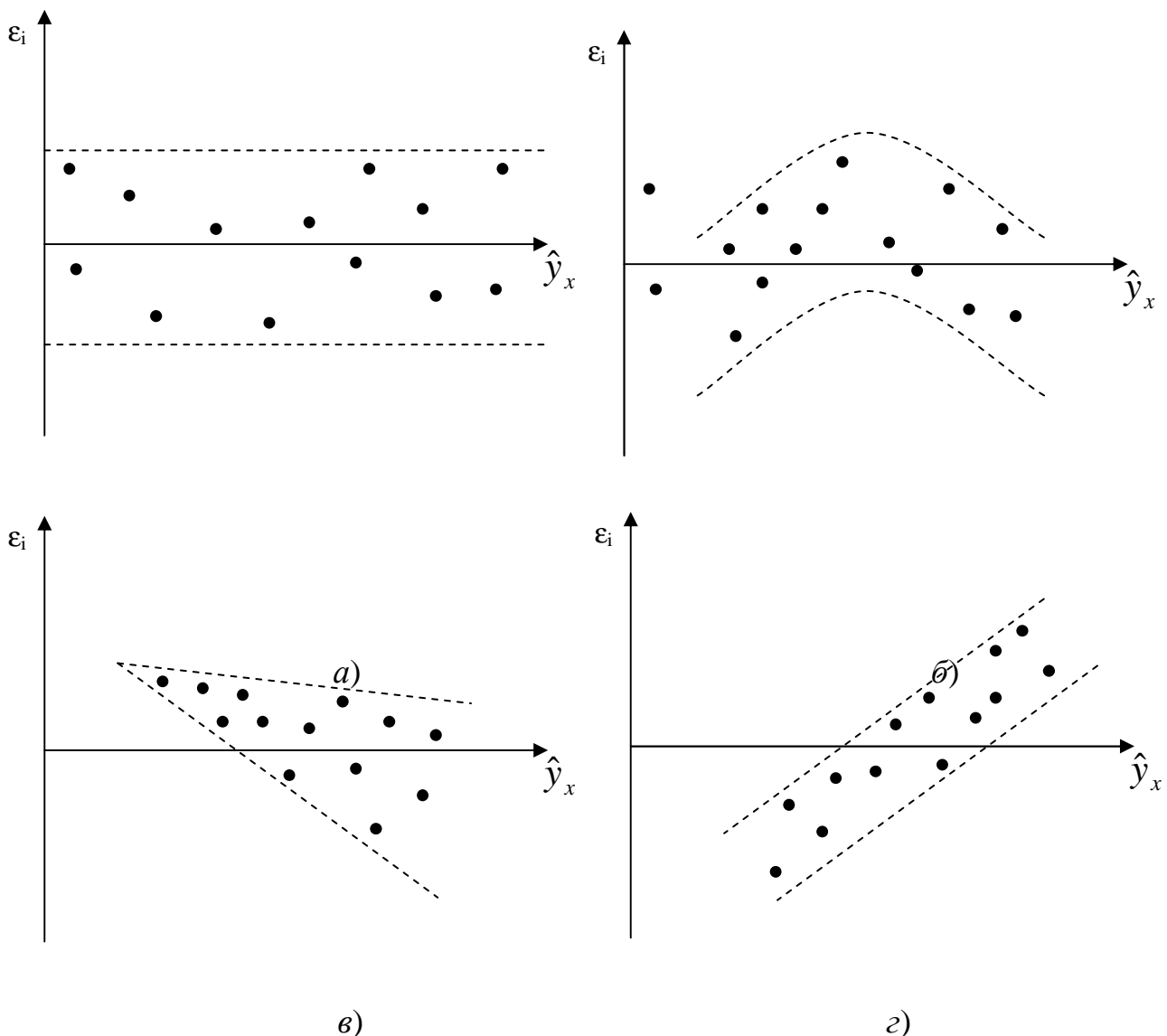


Рисунок 12.1 - Залежність випадкових залишків  $\varepsilon_i$  від теоретичних значень  $\hat{y}_x$  :

*a* - залишки  $\varepsilon_i$  є випадковими величинами; *б* - залишки  $\varepsilon_i$  не випадкові;

*в* - залишки  $\varepsilon_i$  не мають постійної дисперсії;

*г* - залишки  $\varepsilon_i$  носять систематичний характер

Якщо залишки на графіку розташовані у вигляді горизонтальної смуги точок, то вони незалежні від значень  $x_j$ . Якщо ж графік показує наявність залежності  $\varepsilon_i$  і  $x_j$ , то модель неадекватна. Причини неадекватності можуть бути різні. Можливо, що порушено третю передумову МНК, і дисперсія залишків не постійна для кожного значення фактора  $x_j$ . Може бути невірною специфікація моделі, і до неї необхідно ввести додаткові члени від  $x_j$ , наприклад  $x_j^2$ . Скупчення точок у певних ділянках значень фактора  $x_j$  говорить про наявність систематичної помилки моделі.

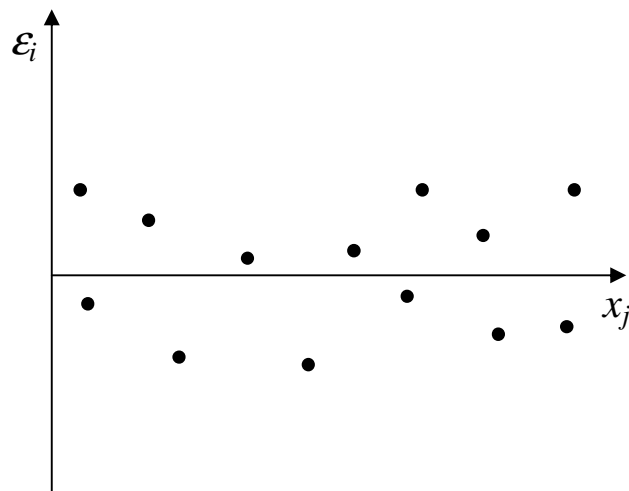


Рисунок 12.2 - Залежність величини залишків  $\varepsilon_i$  від величини фактору  $x_j$

Передумова про нормальний розподіл залишків дозволяє проводити перевірку параметрів регресії за допомогою  $F$ - і  $t$ -критеріїв. Разом з тим, оцінки параметрів регресії, знайдені із застосуванням МНК, мають гарні властивості навіть при відсутності нормального розподілу залишків, тобто при порушенні п'ятої передумови МНК.

Зовсім необхідним для одержання за МНК спроможних оцінок параметрів регресії є дотримання третьої і четвертої передумов.

Відповідно до третьої передумови МНК потрібно, щоб дисперсія залишків була **гомоскедастичною**. Це значить, що для кожного значення фактору  $x_j$  залишки  $\varepsilon_i$  мають однакову дисперсію. Якщо ця умова застосування МНК не виконується, то має місце **гетероскедастичність**. Наявність гетероскедастичності показано на рисунку 12.3, в.

Для множинної регресії є найбільш прийнятний візуальний спосіб вивчення гомо- і гетероскедастичності на підставі графіків залежності залишків  $\varepsilon_i$  від теоретичних значень результативної ознаки  $\hat{y}_x$ .

При побудові регресійних моделей надзвичайно важливе дотримання четвертої передумови МНК, що полягає у відсутності **автокореляції залишків**, тобто значення залишків  $\varepsilon_i$  розподілені незалежно один від одного. Автокореляція залишків означає наявність кореляції між залишками поточних і попередніх (наступних) спостережень. Коефіцієнт кореляції між  $\varepsilon_i$  і  $\varepsilon_j$ , де  $\varepsilon_i$  –

залишки поточних спостережень,  $\varepsilon_j$  – залишки попередніх спостережень (наприклад,  $j=i-1$ ), може бути визначений як

$$r_{\varepsilon_i \varepsilon_j} = \frac{\text{cov}(\varepsilon_i \varepsilon_j)}{\sigma_{\varepsilon_i} \sigma_{\varepsilon_j}},$$

тобто за звичайною формулою лінійного коефіцієнта кореляції. Якщо цей коефіцієнт виявиться істотно відмінним від нуля, то залишки автокорельовані і функція щільності імовірності  $f(\varepsilon)$  залежить від  $j$ -ї точки спостереження і від розподілу значень залишків в інших точках спостереження.

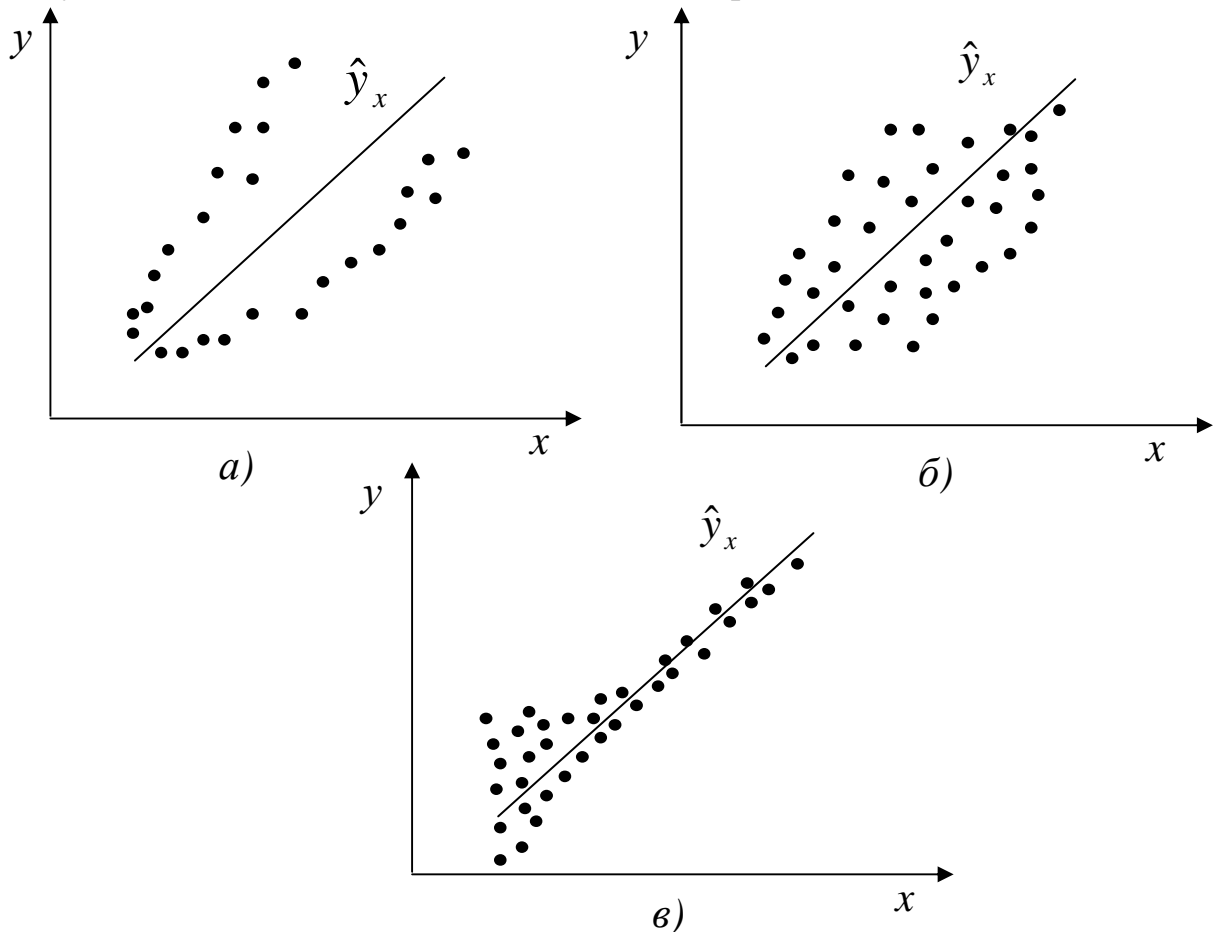


Рисунок 12.3 - Приклади гетероскедастичності:

*а* - дисперсія залишків зростає в міру збільшення  $x$ ; *б* - дисперсія залишків досягає максимальної величини при середніх значеннях змінної  $x$  і зменшується при мінімальних і максимальних значеннях  $x$ ; *в* - максимальна дисперсія залишків при малих значеннях  $x$  і стає однорідною в міру збільшення значень  $x$

Відсутність автокореляції залишкових величин забезпечує спроможність і ефективність оцінок коефіцієнтів регресії. Особливо актуальне дотримання даної передумови МНК при побудові регресійних моделей за рядами динаміки, де у зв'язку з наявністю тенденції наступні рівні динамічного ряду, як правило, залежать від своїх попередніх рівнів.

При недотриманні основних передумов МНК доводиться корегувати модель, змінюючи її специфікацію, додавати або виключати деякі фактори,

перетворювати вихідні дані для того, щоб одержати оцінки коефіцієнтів регресії, які мають властивість незміщеності, мають менше значення дисперсії залишків і забезпечують в зв'язку з цим більш ефективну статистичну перевірку значущості параметрів регресії.

## 12.2 Узагальнений метод найменших квадратів (УМНК)

При порушенні властивості гомоскедастичності й наявності автокореляції залишків  $\varepsilon_i$  рекомендують традиційний метод найменших квадратів замінити узагальненим методом.

Узагальнений метод найменших квадратів застосовують до перетворених даних, він дозволяє одержувати оцінки, які володіють не тільки властивістю незміщеності, але й мають менші вибірккові дисперсії. Зупинимось на використанні УМНК для корегування гетероскедастичності.

Як і раніше, будемо припускати, що середнє значення залишкових величин дорівнює нулю. А от їхня дисперсія не залишається незмінною для різних значень фактора, а пропорційна величині  $K_i$ , тобто

$$\sigma_{\varepsilon_i}^2 = \sigma^2 K_i,$$

де  $\sigma_{\varepsilon_i}^2$  – дисперсія помилки при конкретному  $i$ -му значенні фактору;

$\sigma^2$  – постійна дисперсія помилки при дотриманні передумови про гомоскедастичність залишків;

$K_i$  – коефіцієнт пропорційності, що змінюється зі зміною величини фактора, що й зумовлює неоднорідність дисперсії.

При цьому вважають, що  $\sigma^2$  невідома, а відносно величин  $K_i$  висувують певні гіпотези, що характеризують структуру гетероскедастичності.

У загальному виді для рівняння  $y_i = a + bx_i + \varepsilon_i$  при  $\sigma_{\varepsilon_i}^2 = \sigma^2 K_i$  модель прийме вигляд:  $y_i = a + bx_i + \sqrt{K_i} \varepsilon_i$ . В ній залишкові величини є гетероскедастичними. Припускаючи в них відсутність автокореляції, можна перейти до рівняння з гомоскедастичними залишками, розділивши всі змінні, зафіксовані в ході  $i$ -го спостереження, на  $\sqrt{K_i}$ . Тоді дисперсія залишків буде величиною постійною, тобто  $\sigma_{\varepsilon_i}^2 = \sigma^2$ .

Іншими словами, від регресії  $y$  на  $x$  ми перейдемо до регресії на нові змінні:  $\frac{y}{\sqrt{K}}$  і  $\frac{x}{\sqrt{K}}$ . Рівняння регресії прийме вигляд:

$$\frac{y_i}{\sqrt{K_i}} = \frac{a}{\sqrt{K_i}} + b \frac{x_i}{\sqrt{K_i}} + \varepsilon_i,$$

а вихідні дані для цього рівняння матимуть вигляд:

$$y = \begin{pmatrix} \frac{y_1}{\sqrt{K_1}} \\ \frac{y_2}{\sqrt{K_2}} \\ \dots \\ \frac{y_n}{\sqrt{K_n}} \end{pmatrix}, \quad x = \begin{pmatrix} \frac{x_1}{\sqrt{K_1}} \\ \frac{x_2}{\sqrt{K_2}} \\ \dots \\ \frac{x_n}{\sqrt{K_n}} \end{pmatrix}.$$

Стосовно звичайної регресії рівняння з новими перетвореними змінними являє собою зважену регресію, в якій змінні  $y$  і  $x$  взяті з вагами  $\frac{1}{\sqrt{K}}$ .

Оцінка параметрів нового рівняння з перетвореними змінними приводить до зваженого методу найменших квадратів, для якого необхідно мінімізувати суму квадратів відхилень виду

$$S(a,b) = \sum_{i=1}^n \frac{1}{K_i} (y_i - a - bx_i)^2.$$

Відповідно дістанемо наступну систему нормальних рівнянь:

$$\begin{cases} \sum_{i=1}^n \frac{y_i}{K_i} = a \sum_{i=1}^n \frac{1}{K_i} + b \sum_{i=1}^n \frac{x_i}{K_i}, \\ \sum_{i=1}^n \frac{y_i x_i}{K_i} = a \sum_{i=1}^n \frac{x_i}{K_i} + b \sum_{i=1}^n \frac{x_i^2}{K_i}. \end{cases}$$

Якщо перетворені змінні  $x$  і  $y$  взяті у відхиленнях від середніх рівнів, то коефіцієнт регресії  $b$  можна визначити як

$$b = \frac{\sum_{i=1}^n \frac{1}{K_i} x_i y_i}{\sum_{i=1}^n \frac{1}{K_i} x_i^2}.$$

При звичайному застосуванні методу найменших квадратів до рівняння лінійної регресії для змінних у відхиленнях від середніх рівнів коефіцієнт регресії  $b$  визначають за формулою:

$$b = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

Як бачимо, при використанні узагальненого МНК з метою корегування гетероскедастичності коефіцієнт регресії  $b$  являє собою зважену величину стосовно звичайного МНК з вагою  $\frac{1}{\sqrt{K}}$ .

Аналогічний підхід можливий не тільки для рівняння парної, але і для множинної регресії. Припустимо, що розглядають модель виду

$$y = a + b_1 x_1 + b_2 x_2 + \varepsilon,$$

для якої дисперсія залишкових величин виявилася пропорційною  $K_i^2$ .  $K_i$  являє собою коефіцієнт пропорційності, що приймає різні значення для відповідних і значень факторів  $x_1$  і  $x_2$ . Через те, що  $\sigma_{\varepsilon_i}^2 = \sigma^2 K_i^2$ , розглянута модель прийме вигляд

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + K_i \varepsilon_i,$$

де залишки  $\varepsilon_i$  є гетероскедастичними.

Для того щоб одержати рівняння, де залишки  $\varepsilon_i$  є гомоскедастичними, слід перейти до нових перетворених змінних, розділивши всі члени вихідного рівняння на коефіцієнт пропорційності  $K$ . Рівняння з перетвореними змінними матиме вигляд

$$\frac{y_i}{K_i} = A + b_1 \frac{x_{1i}}{K_i} + b_2 \frac{x_{2i}}{K_i} + \varepsilon_i.$$

Це рівняння не містить вільного члена. Разом з тим, знайшовши змінні в новому перетвореному вигляді і застосовуючи звичайний МНК до них, одержимо іншу специфікацію моделі:

$$\frac{y_i}{K_i} = A + b_1 \frac{x_{1i}}{K_i} + b_2 \frac{x_{2i}}{K_i} + \varepsilon_i.$$

Параметри такої моделі залежать від концепції, прийнятої для коефіцієнта пропорційності  $K_i$ . В економетричних дослідженнях досить часто пропонують гіпотезу, що залишки  $\varepsilon_i$  пропорційні значенням фактора. Так, якщо в рівнянні

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m + e$$

припустити, що  $e = \varepsilon x_1$ , тобто  $K = x_1$  і  $\sigma_{\varepsilon_i}^2 = \sigma^2 x_1$ , то узагальнений МНК припускає оцінку параметрів наступного трансформованого рівняння:

$$\frac{y}{x_1} = b_1 + b_2 \frac{x_2}{x_1} + \dots + b_m \frac{x_m}{x_1} + \varepsilon.$$

Застосування в цьому випадку узагальненого МНК приводить до того, що спостереження з меншими значеннями перетворених змінних  $\frac{x}{K}$  мають при визначенні параметрів регресії відносно більшу вагу, ніж з первісними змінними. Разом з тим, треба мати на увазі, що нові перетворені змінні одержують новий економічний зміст і їхня регресія має інший зміст, ніж регресія за вихідними даними.

Розглянемо приклад. Нехай  $y$  – витрати виробництва,  $x_1$  – обсяг продукції,  $x_2$  – основні виробничі фонди,  $x_3$  – чисельність працівників, тоді рівняння

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + e$$

є моделлю витрат виробництва з об'ємними факторами. Припускаючи, що  $\sigma_{\varepsilon_i}^2$  пропорційна квадрату чисельності працівників  $x_3$ , ми одержимо як результативну ознаку витрати на одного працівника  $\frac{y}{x_3}$ , а як фактори –

наступні показники: продуктивність праці  $\frac{x_1}{x_3}$  і фондоозброєність праці  $\frac{x_2}{x_3}$ .

Відповідно трансформована модель матиме вигляд

$$\frac{y}{x_3} = b_3 + b_1 \frac{x_1}{x_3} + b_2 \frac{x_2}{x_3} + \varepsilon,$$

де параметри  $b_1, b_2, b_3$  чисельно не збігаються з аналогічними параметрами попередньої моделі. Крім цього, коефіцієнти регресії змінюють економічний зміст: з показників сили зв'язку, що характеризують середню абсолютну зміну витрат виробництва зі зміною абсолютної величини відповідного фактора на одиницю, вони фіксують при узагальненому МНК середню зміну витрат на працівника; зі зміною продуктивності праці на одиницю при незмінному рівні фондоозброєності праці; і зі зміною фондоозброєності праці на одиницю при незмінному рівні продуктивності праці.

Якщо припустити, що в моделі з первісними змінними дисперсія залишків пропорційна квадрату обсягу продукції,  $\sigma_{\varepsilon_i}^2 = \sigma^2 x_1^2$ , можна перейти до рівняння регресії виду

$$\frac{y}{x_1} = b_1 + b_2 \frac{x_2}{x_1} + b_3 \frac{x_3}{x_1} + \varepsilon.$$

У ньому нові змінні:  $\frac{y}{x_1}$  – витрати на одиницю (або на 1 грн. продукції),

$\frac{x_2}{x_1}$  – фондомісткість продукції,  $\frac{x_3}{x_1}$  – трудомісткість продукції.

Гіпотеза про те, що залишки пропорційні величині фактора, може мати реальну підставу: при обробці недостатньо однорідної сукупності, що включає як великі, так і малі підприємства, більшим об'ємним значенням фактора можуть відповідати більші дисперсії результативної ознаки і дисперсії залишкових величин.

За наявності однієї пояснюючої змінної гіпотеза  $\sigma_{\varepsilon_i}^2 = \sigma^2 x^2$  трансформує лінійне рівняння

$$y = a + bx + e$$

на рівняння

$$\frac{y}{x} = b + \frac{a}{x} + \varepsilon,$$

в якому параметри  $a$  і  $b$  помінялися місцями, константа стала коефіцієнтом нахилу лінії регресії, а коефіцієнт регресії – вільним членом.



Перехід до відносних величин істотно знижує варіацію фактора і відповідно зменшує дисперсію помилки. Він являє собою найбільш простий випадок обліку гетероскедастичності в регресійних моделях за допомогою узагальненого МНК. Процес переходу до відносних величин може бути ускладнений висуванням інших гіпотез про пропорційність помилок щодо включених до моделі факторів. Використання тієї або іншої гіпотези припускає спеціальні дослідження залишкових величин для відповідних регресійних моделей. Застосування узагальненого МНК дозволяє одержати оцінки параметрів моделі, що володіють меншою дисперсією.

### Контрольні запитання

1. Якими властивостями мають володіти оцінки параметрів регресії?
2. Як впливає на параметри множинної лінійної моделі порушення умови, що математичне сподівання збурення  $\varepsilon$  дорівнює нулю?
3. Як впливає на параметри множинної лінійної моделі порушення умови, що дисперсія збурення  $\varepsilon$  постійною?
4. Який спосіб використовують для вивчення гомо- і гетероскедастичності?
5. На підставі якого дослідження роблять висновок про автокорельованість залишків?
6. Як поведуться при недотриманні основних передумов МНК?
7. Чим відрізняється узагальнений МНК від звичайного? У яких випадках його використовують?
8. Які властивості притаманні оцінкам параметрів моделі, отриманим на основі узагальненого МНК?

## ТЕМА 13 ЕКОНОМЕТРИЧНІ МОДЕЛІ ДИНАМІКИ

### 13.1 Загальні відомості про часові ряди і завдання їхнього аналізу

При розгляді класичної моделі регресії характер експериментальних даних, як правило, не має принципового значення. Однак це виявляється не так, коли порушені умови класичної моделі. Методи дослідження моделей, заснованих на даних просторових вибірок і часових рядів істотно відрізняються. Пояснюється це тим, що на відміну від просторових вибірок спостереження в часових рядах, як правило, не можна вважати незалежними.

**Часовий ряд** (ряд динаміки) – це сукупність значень будь-якого показника за кілька послідовних моментів або періодів часу. Кожний рівень часового ряду формується під впливом великої кількості факторів, які умовно можна підрозділити на три групи:

- фактори, що формують тенденцію ряду;

- фактори, що формують циклічні коливання ряду;
- випадкові фактори.

Більшість часових рядів економічних показників мають **тенденцію**, що характеризує сукупний довгостроковий вплив множини факторів на динаміку досліджуваного показника. Всі ці фактори, взяті окремо, можуть робити різнонаправлений вплив на досліджуваний показник. Однак у сукупності вони формують його зростаючу або спадну тенденцію. На рисунку 13.1 показаний гіпотетичний часовий ряд, що містить зростаючу тенденцію.

Досліджуваний показник може також бути підданий циклічним коливанням. Ці коливання можуть носити сезонний характер, оскільки економічна діяльність ряду галузей економіки залежить від пори року. На рисунку 13.2 поданий гіпотетичний часовий ряд, що містить тільки сезонну компоненту.

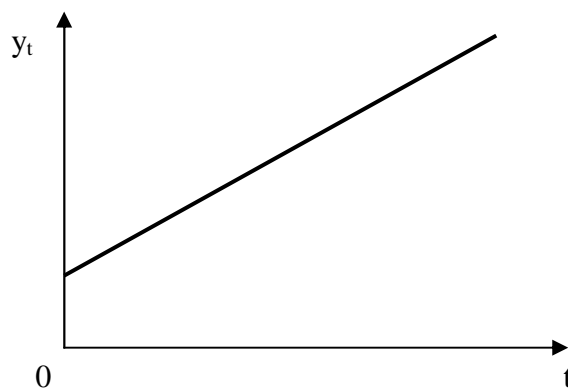


Рисунок 13.1 - Часовий ряд, що містить зростаючу тенденцію

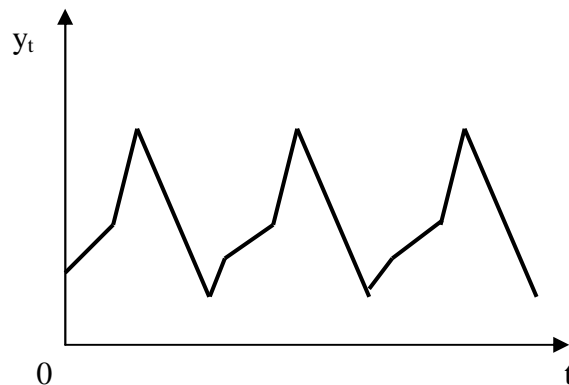


Рисунок 13.2 - Часовий ряд, що містить тільки сезонну компоненту

Деякі часові ряди не містять тенденції і циклічної компоненти, а кожний наступний їхній рівень утворюється як сума середнього рівня ряду і якоїсь (додатної або від'ємної) випадкової компоненти. Приклад ряду, що містить тільки випадкову компоненту, наведений на рисунку 13.3.

Очевидно, що реальні дані не впливають цілком з будь-яких описаних вище моделей. Найчастіше вони містять всі три компоненти. Кожний їхній рівень формується під впливом тенденції, сезонних коливань і випадкової компоненти.

У більшості випадків фактичний рівень часового ряду можна представити як суму або добуток трендової, циклічної і випадкової компонентів. Модель, в якій часовий ряд представлений як сума перелічених компонентів, називається **адитивною моделлю** часового ряду. Модель, в якій часовий ряд представлений як добуток перелічених компонентів, називається **мультиплікативною моделлю** часового ряду. Основна задача економетричного дослідження окремого часового ряду - виявлення і побудова кількісного виразу кожної з перелічених вище компонентів для того, щоб використати отриману інформацію для прогнозування майбутніх значень ряду або при побудові моделей взаємозв'язку двох або більше часових рядів.

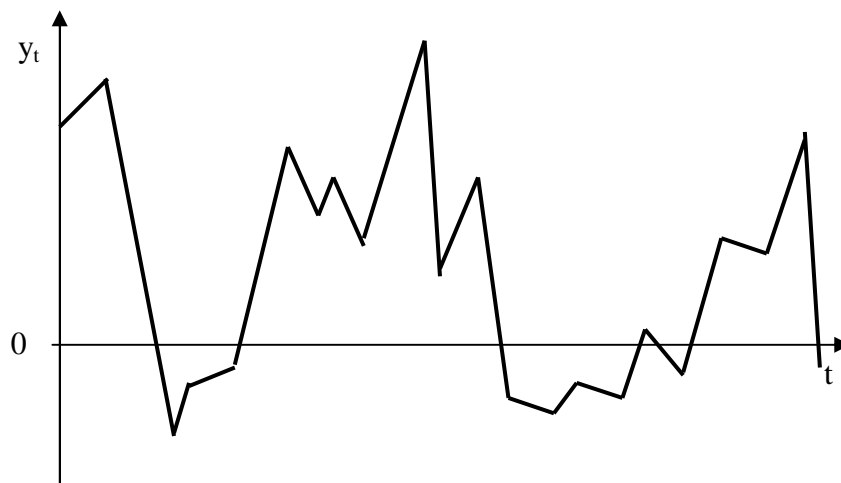


Рисунок 13.3 - Часовий ряд, що містить тільки випадкову компоненту

Найважливішою класичною задачею при дослідженні економічних часових рядів є виявлення і статистична оцінка основної тенденції розвитку досліджуваного процесу і відхилень від неї, або **тренду**.

Тренд або тенденція часового ряду - це декілька умовне поняття. Під трендом розуміють закономірну, не випадкову складову часового ряду (звичайно монотонну), що може бути обчисленою за цілком визначеним однозначним правилом. Тренд часового ряду часто пов'язаний з дією фізичних законів або яких-небудь інших об'єктивних закономірностей. Однак, загалом кажучи, не можна однозначно розділити випадковий процес або часовий ряд на регулярну частину (тренд) і коливальну частину (залишок). Тому звичайно припускають, що тренд - це деяка функція простого виду (лінійна, квадратична і та ін.), що описує "поведінку в цілому" ряду або процесу. Якщо виділення такого тренда спрощує дослідження, то припущення про обрану форму тренда вважається припустимим.

Після виділення лінійного тренда потрібно з'ясувати, наскільки він значущий. Це робиться за допомогою аналізу коефіцієнту кореляції. Справа в тому, що відмінність коефіцієнта кореляції від нуля і тим самим наявність реального тренда (позитивного або негативного) може виявитися випадковою, зв'язаною із специфікою розглянутого відрізка часового ряду. Інакше кажучи, при аналізі іншого набору експериментальних даних (для того ж часового ряду) може виявитися, що отримана при цьому оцінка набагато ближче до нуля, ніж

вихідна (і, можливо, навіть має інший знак), і говорити про реальний тренд тут уже стає важко. Такі ряди називаються часовими рядами з детерміністичним трендом. Існують також часові ряди із стохастичним трендом.

Відзначимо основні етапи аналізу часових рядів:

- графічне подання і опис поведінки часового ряду;
- виділення і видалення закономірних (невипадкових) складових часового ряду (тренду, сезонних і циклічних складових);
- згладжування і фільтрація (видалення низько- або високочастотних складових часового ряду);
- дослідження випадкової складової часового ряду, побудова і перевірка адекватності математичної моделі для його опису;
- прогнозування розвитку досліджуваного процесу на основі наявного часового ряду;
- дослідження взаємозв'язку між різними часовими рядами.

Серед найпоширеніших методів аналізу часових рядів є моделі авторегресії і ковзної середньої, кореляційний і спектральний аналіз.

Якщо вибірку  $y_1, y_2, \dots, y_n$  розглядають як одну з реалізацій випадкової величини  $Y$ , часовий ряд  $y_1, y_2, \dots, y_n$  розглядається як одна з реалізацій (траєкторій) випадкового процесу  $Y(t)$ . Разом з тим необхідно мати на увазі принцип відмінності часового ряду  $y_t$ ,  $(t=1, n)$  від випадкової вибірки. По-перше, на відміну від елементів випадкової вибірки члени часового ряду, як правило, не є статистично незалежними. По-друге, члени часового ряду не є однаково розподіленими.

За наявності в часовому ряду тенденції і циклічних коливань значення кожного наступного рівня ряду залежать від попередніх. Кореляційну залежність між послідовними рівнями часового ряду називають **автокореляцією рівнів ряду**. Кількісно її можна виміряти за допомогою лінійного коефіцієнта кореляції між рівнями часового ряду, зрушеними на кілька кроків у часі. Формула для розрахунку коефіцієнта автокореляції має вигляд:

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}}, \quad (13.1)$$

де  $\bar{y}_1 = \frac{1}{n-1} \sum_{t=2}^n y_t$ ,  $\bar{y}_2 = \frac{1}{n-1} \sum_{t=2}^n y_{t-1}$ ,

Цю величину називають коефіцієнтом автокореляції рівнів ряду першого порядку, тому що він вимірює залежність між сусідніми рівнями ряду  $y_t$  і  $y_{t-1}$ .

Аналогічно можна визначити коефіцієнти автокореляції другого і більш високих порядків. Так, коефіцієнт автокореляції другого порядку характеризує тісноту зв'язку між рівнями  $y_t$  і  $y_{t-2}$  і визначають його за формулою:

$$r_1 = \frac{\sum_{t=3}^n (y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^n (y_t - \bar{y}_3)^2 \sum_{t=3}^n (y_{t-2} - \bar{y}_4)^2}}, \quad (13.2)$$

$$\text{де } \bar{y}_3 = \frac{1}{n-2} \sum_{t=3}^n y_t, \quad \bar{y}_4 = \frac{1}{n-2} \sum_{t=3}^n y_{t-2}.$$

Число періодів, за якими розраховується коефіцієнт автокореляції, називають **лагом**. Із збільшенням лагу число пар значень, за якими розраховують коефіцієнт автокореляції, зменшується. Вважають за доцільне для забезпечення статистичної вірогідності коефіцієнтів автокореляції

використовувати правило – максимальний лаг не повинен перевищувати  $\tau \leq \frac{n}{4}$ .

Перелічимо властивості коефіцієнта автокореляції.

1. Його будують за аналогією з лінійним коефіцієнтом кореляції і в такий спосіб він характеризує тісноту тільки лінійного зв'язку поточного і попереднього рівнів ряду. Тому з коефіцієнта автокореляції можна судити про наявність лінійної (або близької до лінійної) тенденції. Для деяких часових рядів, що мають сильну нелінійну тенденцію (наприклад, параболу другого порядку або експоненту), коефіцієнт автокореляції рівнів вихідного ряду може наближатися до нуля.

2. За знаком коефіцієнта автокореляції не можна робити висновок про зростаючу або спадну тенденцію в рівнях ряду. Більшість часових рядів економічних даних містять додатну автокореляцію рівнів, однак при цьому можуть мати спадну тенденцію.

Послідовність коефіцієнтів автокореляції рівнів першого, другого і т.д. порядків називають **автокореляційною функцією** часового ряду. Графік залежності її значень від величини лагу (порядку коефіцієнта автокореляції) називають **корелограмою**.

Аналіз автокореляційної функції і корелограми дозволяє визначити лаг, при якому автокореляція є найбільш високою, а отже, і лаг, при якому зв'язок між поточним і попереднім рівнями ряду найбільш тісний, тобто за допомогою аналізу автокореляційної функції і корелограми можна виявити структуру ряду.

Якщо найбільш високим виявився коефіцієнт автокореляції першого порядку, досліджуваний ряд містить тільки тенденцію. Якщо найбільш високим виявився коефіцієнт автокореляції порядку  $\tau$ , то ряд містить циклічні коливання з періодичністю  $\tau$  моментів часу. Якщо жоден з коефіцієнтів автокореляції не є значущим, можна зробити одне з двох припущень щодо структури цього ряду: або ряд не містить тенденції і циклічних коливань, або ряд містить сильну нелінійну тенденцію, для виявлення якої потрібно провести додатковий аналіз. Тому коефіцієнт автокореляції рівнів і автокореляційну функцію доцільно використовувати для виявлення в часовому ряді наявності або відсутності трендової компоненти і циклічної (сезонної) компоненти.

### 13.2 Моделювання тенденції часового ряду

Розповсюдженим способом моделювання тенденції часового ряду є побудова аналітичної функції, що характеризує залежність рівнів ряду від часу, або тренду. Цей спосіб називають аналітичним вирівнюванням часового ряду.

Оскільки залежність від часу може приймати різні форми, для її формалізації можна використати різні види функцій. Для побудови трендів найчастіше застосовують функції, що наведені у таблиці 13.1.

Таблиця 13.1 – Деякі функції для побудови трендів

Вид рівняння	Відбивана рівнянням тенденція
Рівняння прямої $\bar{y}_t = at + b$	Рівномірне зростання при $a > 0$ або рівномірне падіння при $a < 0$
Експонентна функція $\bar{y}_t = a^t b$	Прискорюване зростання при $a > 1$ або падіння, що вповільнюється, при $a < 1$
Гіпербола $\bar{y}_t = \frac{a}{t} + b$	Вповільнюване падіння, при $a > 0$ або вповільнюване зростання, при $a < 0$
Парабола $\bar{y}_t = at^2 + bt + c$	Зростання, що переходить у падіння, або падіння, що переходить у зростання у точці $t = -\frac{b}{2a}$

Параметри наведених трендів можна визначити за звичайним методом найменших квадратів (МНК), використовуючи як незалежну змінну час  $t=1, n$ , а як залежну змінну – фактичні рівні часового ряду  $\hat{y}_t$ .

$$y_t = f(t) + \varepsilon_i, \quad (13.3)$$

де  $\varepsilon_i$  – збурення, що задовольняють основним передумовам регресійного аналізу, тобто уявляють собою незалежні і однаково розподілені випадкові величини, розподіл яких передбачається нормальним.

Нагадаємо, що відповідно до МНК параметри лінійного тренду  $\hat{y}_t = f(t) = b_0 + b_1 t$  знаходять з системи нормальних рівнянь, в якій у якості пояснюючої змінної  $x_i$  фігурує час  $t$ .

$$\begin{cases} b_0 n + b_1 \sum_{t=1}^n t = \sum_{i=1}^n y_i, \\ b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2 = \sum_{i=1}^n y_i t. \end{cases} \quad (13.4)$$

При застосуванні МНК для оцінки параметрів експонентної або логістичної функцій виникають складності з розв'язанням системи нормальних рівнянь, тому попередньо, до одержання відповідної системи, звертаються до

перетворення цих функцій. Для нелінійних трендів попередньо проводять стандартну процедуру їхньої лінеаризації.

Іншим методом вирівнювання (згладжування) часового ряду, тобто виділення не випадкової складової, є метод **ковзних середніх**. Він заснований на переході від початкових значень членів ряду до їхніх середніх значень на інтервалі часу, довжина якого визначена заздалегідь. При цьому сам обраний інтервал часу «ковзає» уздовж ряду.

Одержаний у такий спосіб ряд ковзних середніх поводить себе більш гладко, ніж вихідний ряд, через усереднення відхилень ряду. Дійсно, якщо індивідуальний розкид значень члена часового ряду  $y_i$ , біля свого середнього (згладженого) значення характеризується дисперсією  $\sigma^2$ , то розкид середньої з  $m$  членів часового ряду  $(y_1 + y_2 + \dots + y_m)/m$  біля того самого значення буде  $\frac{\sigma^2}{m}$ .

характеризуватися істотно меншою величиною дисперсії, що дорівнює  $\frac{\sigma^2}{m}$ . Для усереднення може бути використана середня арифметична (проста й з деякими вагами), медіана та ін.

Існує декілька способів визначення типу тенденції. До числа найпоширеніших способів відносяться якісний аналіз досліджуваного процесу, побудова і візуальний аналіз графіка залежності рівнів ряду від часу. З цією ж метою можна використати і коефіцієнти автокореляції рівнів ряду. Тип тенденції можна визначити шляхом порівняння коефіцієнтів автокореляції першого порядку, розрахованих за вихідними і перетвореними рівнями ряду.

Якщо часовий ряд має лінійну тенденцію, то його сусідні рівні  $\hat{y}_t$  і  $\hat{y}_{t-1}$  тісно корелюють. В цьому випадку коефіцієнт автокореляції першого порядку рівнів вихідного ряду повинен бути високим. Якщо часовий ряд містить нелінійну тенденцію, наприклад, у формі експоненти, то коефіцієнт автокореляції першого порядку за логарифмами рівнів вихідного ряду буде вище, ніж відповідний коефіцієнт, розрахований за рівнями ряду. Чим чіткіше виражена нелінійна тенденція в досліджуваному часовому ряді, тим більшою мірою будуть розрізнятися значення зазначених коефіцієнтів.

Вибір найкращого рівняння у випадку, коли ряд містить нелінійну тенденцію, можна здійснити шляхом перебору основних форм тренду, розрахунку за кожним рівнянням скорегованого коефіцієнта детермінації і середньої помилки апроксимації. Цей метод легко реалізують при комп'ютерній обробці даних.

### 13.3 Аналіз аддитивної і мультиплікативної моделей часового ряду

Аналіз часового ряду полягає у виділенні окремих компонент. Методика аналізу залежить від того, який зв'язок між цими компонентами: аддитивний або мультиплікативний.

**Аддитивною моделлю** часового ряду називається така модель, у якій зміна значень змінної в часі описується через додавання окремих компонентів:

$$Y = T + S + E. \quad (13.5)$$

Процедура аналізу аддитивного часового ряду включає:

- розрахунок значень сезонної компоненти  $S$ ;
- обчислення сезонної, компоненти з фактичних значень (десезоналізація даних), тобто  $Y-S$ ;
- розрахунок тренду  $T$  на основі десезоналізованих даних;
- розрахунок помилок як різниць між фактичними й трендовими значеннями  $E = Y - T$ ;
- розрахунок помилки апроксимації - середнього відхилення або MAD або середньоквадратичної помилки MSE.

Для виділення сезонної компоненти провадиться усунення сезонних коливань за методом ковзної середньої. На основі рівнів часового ряду розраховуються ковзні середні, які звільнені від сезонних коливань, але включають випадкову компоненту.

Наприклад, для виключення сезонних коливань з поквартальних даних знаходять усереднені «ковзні» рівні:

$$\bar{y}_3 = \frac{1/2y_1 + y_2 + y_3 + y_4 + 1/2y_5}{4}, \quad (13.6)$$

$$\bar{y}_4 = \frac{1/2y_2 + y_3 + y_4 + y_5 + 1/2y_6}{4}, \quad (13.7)$$

$$\bar{y}_5 = \frac{1/2y_3 + y_4 + y_5 + y_6 + 1/2y_7}{4} \text{ і т.д.}, \quad (13.8)$$

де  $\bar{y}_i = T_i + E_i$  - елемент часового ряду, що містить тренд і випадкову компоненту.

Тоді виділення сезонної компоненти провадиться на підставі рівності

$$y - \bar{y} = S + E. \quad (13.9)$$

Сезонна компонента знаходиться як середнє значення сезонних оцінок для кожного сезону незалежно від особливостей року. Загальна сума сезонних оцінок повинна дорівнювати нулю:  $\sum S_j = 0$  (тут  $j$  - номер сезону). Це необхідно, щоб усереднити значення сезонної компоненти в цілому за рік. Тому отримані сезонні оцінки доводиться коректувати, щоб виконати цю умову. Після цього усувають сезонну компоненту з фактичних даних, тобто знаходять  $Y - S = T + E$ . Дані такого роду називаються десезоналізованими. Їх використовують для побудови рівняння тренду.

Рівняння лінійного тренду має вигляд:

$$T = a + bt,$$

де  $t$  - номер кварталу (або місяця);

$a$  - відрізок, що відтинається лінією тренду при перетинанні з віссю ординат;

$b$  - характеристика нахилу лінії тренду до осі абсцис.

Параметри  $a$  і  $b$  визначають за методом найменших квадратів. Рівняння для розрахунку параметрів  $a$  і  $b$  мають вигляд:

$$b = \frac{n \sum ty - \sum t \sum y}{n \sum t^2 - (\sum t)^2}; \quad a = \frac{\sum y}{n} - \frac{b \sum t}{n},$$

де  $t$  - порядковий номер кварталу;



$y = T + E$  - десеоналізовані рівні часового ряду.

Знайшовши сезонну компоненту і тренд, виділяють випадкову компоненту:

$$Y - S - T = E. \quad (13.10)$$

Значення випадкової компоненти (помилки) використовують для розрахунку середнього абсолютного відхилення:

$$MAD = \frac{\sum_{i=1}^n |E_i|}{n}, \quad (13.11)$$

або середньоквадратичної помилки

$$MSE = \frac{\sum_{i=1}^n E_i^2}{n}. \quad (13.12)$$

Якщо помилки малі, то роблять висновок про те, що тенденція стала і дозволяє одержати гарні короткострокові прогнози. Прогнозні значення за аддитивною моделлю розраховують так

$$Y = T + S, \quad (13.13)$$

де  $T$  - трендове значення для відповідного кварталу (місяця);

$S$  - сезонна компонента для відповідного кварталу (місяця).

Трендове значення для прогнозного кварталу (місяця) розраховують за рівнянням тренду:

$$T = a + bt, \quad (13.14)$$

де  $t$  - порядковий номер прогнозного кварталу (місяця).

Чим менше період попередження, тим більш обґрунтованим є прогноз.

Аддитивна модель застосовується при сталості сезонної компоненти. Якщо вона зростає із зростанням тренду, то кращий результат буде отриманий на базі мультиплікативної моделі:

$$Y = T * S * E, \quad (13.15)$$

де  $T$  - трендова компонента;

$S$  - коефіцієнт сезонної компоненти;

$E$  - відносний вплив випадкової компоненти.

У цьому випадку роблять вирівнювання ряду за методом ковзної середньої й знаходять коефіцієнт сезонності:

$$Y Y = S E. \quad (13.16)$$

Якщо часовий ряд побудований за квартальними даними, то сума коефіцієнтів сезонності повинна дорівнювати 4, а якщо за місячними, - то  $\sum_{i=1}^n S_i = 12$ . Інакше корегують коефіцієнти сезонності, щоб виконати цю умову.

На основі десеоналізованих даних розраховують рівняння тренду - лінійне або нелінійне - і знаходять трендові значення  $T$ . Потім обчислюють помилки:

відносну  $E = Y : (T * S);$

абсолютну  $E_a = Y - (T * S).$

Близькість моделі до фактичних даних оцінюється за допомогою показників  $MAD$  і  $MSE$ .

Прогнозні значення визначаються як

$$Y = T * S, \quad (13.17)$$

де  $T$  - прогнозне значення, знайдене з рівняння тренду.

Наприклад,  $T = a + bt$ , де  $t$  - порядковий номер прогнозного кварталу (місяця).

Потім розраховане значення  $T$  коректують на сезонну компоненту відповідного кварталу або місяця:  $T'' \cdot S$ .

### 13.4 Спектральний аналіз часового ряду

При спектральному аналізі виходять з припущення, що часовий ряд є сумою, або спектром багатьох хвилеподібних змін, які можна описати за допомогою тригонометричних функцій. Метою спектрального аналізу є відшукування прихованих періодичностей і оцінка їхньої інтенсивності.

Для опису хвилеподібних коливань динамічного ряду використовують періодичну функцію Фур'є наступного вигляду:

$$\bar{y}_t = a_0 + \sum \left( a_k \cos \frac{360kt}{n} + b_k \sin \frac{360kt}{n} \right),$$

де  $a_0$  - середній рівень ряду;

$k$  - номер гармоніки;

$i$  - порядковий номер часового періоду;

$n$  - довжина ряду, тобто число рівнів у ньому.

Номер гармоніки - це те число хвиль даної довжини, які зможуть укластися в даному ряді. Наприклад, для ряду довжиною в 10 років у хвилі довжиною в 5 років номер гармоніки дорівнюватиме 2. У хвилі довжиною у два роки гармоніка дорівнюватиме 5.

Вищенаведена формула періодичної функції добре відповідає особливостям аналізу періодичних коливань в акустиці, механіці, електротехніці й в інших галузях фізики. Однак вона не підходить для економічних досліджень, де на перше місце висувається не частота коливань, а довжина хвилі  $l$ .

З огляду на те, що  $l = \frac{n}{k}$ , можна надати періодичній функції більш зручний для застосування в економічній галузі вигляд, а саме:

$$\bar{y}_t = a_0 + \sum \left( a_l \cos \frac{360t}{l} + b_l \sin \frac{360t}{l} \right). \quad (13.18)$$

Для знаходження параметрів (13.18) використовують формули:

$$a_0 = \frac{\sum y}{n}; \quad a_l = \frac{2}{n} \sum y \cos \frac{360t}{l}; \quad b_l = \frac{2}{n} \sum y \sin \frac{360t}{l}. \quad (13.19)$$

При цьому дисперсію (амплітуду) хвилі  $l$  можна знайти за формулою:

$$\sigma_l^2 = \frac{a_l^2 + b_l^2}{2}, \quad (13.20)$$

а загальну дисперсію – за формулою:

$$\sigma_{\text{обиц}}^2 = \frac{\sum (y - \bar{y})^2}{n}. \quad (13.21)$$

Частка поясненої варіації ряду, або коефіцієнт детермінації, при цьому буде обчислюватися в такий спосіб:

$$R^2 = \frac{\sum \sigma_l^2}{\sigma_{\text{обиц}}^2}. \quad (13.22)$$

Крім розрахунку коефіцієнтів автокореляції і складання корелограми для виявлення прихованих періодичностей використовують також періодограм-аналіз. Він дозволяє з певною ймовірністю визначити, скільки і якої довжини хвилі (або скільки гармонік і яких номерів) містить аналізований динамічний ряд. Виявлені хвилі можуть мати різну силу. Одні можуть бути дуже слабкими, а інші - сильними. Співвідношення сили окремих хвиль або гармонік показують наочно за допомогою графіка лінійчатого спектра Фур'є. На цьому графіку за абсцисою показують довжини хвиль, а за ординатою їхню потужність. Як міру потужності цих хвиль беруть викликувані ними дисперсії, тобто

$$\sigma_l^2 = \frac{a_l^2 + b_l^2}{2}. \quad (13.23)$$

Якщо всі дисперсії практично рівні, то кажуть, що ряд характеризується «білим шумом». Наявність будь-якої періодичності в зміні рівнів ряду при цьому заперечується.

### 13.5 Прогнозування часового ряду

Будь-який прогноз ґрунтується на перенесенні минулих тенденцій, виявлених у ході аналізу часового ряду, на майбутнє.

Період часу, для якого робиться прогноз, називається горизонтом прогнозування. Відстань до нього від поточного періоду називається періодом попередження. Якщо період дорівнює 1-3 крокам у майбутнє, то кажуть про короткостроковий прогноз. При 4-10 кроках прогноз називають середньостроковим. При більш довгих періодах попередження говорять про довгострокові прогнози.

У ході прогнозування доводиться зустрічатися з наступним протиріччям. Довгий вихідний ряд дозволяє добре погасити всякі випадкові сплески й падіння, але створює небезпеку перенесення на майбутнє занадто старих закономірностей. Короткий ряд виключає таку небезпеку, але не захищає від впливу випадків. Прогноз не може вважатися якісним, якщо на ньому позначилися випадковості або дуже старі закономірності.

Щоб розв'язати дане протиріччя, шукають компроміс між прагненням погасити випадковості й не допустити перенесення на майбутнє занадто старих закономірностей. В умовах різких змін, характерних для нашої країни, знайти такий компроміс дуже важко: занадто короткі ряди треба використовувати, щоб не допустити присутність у прогнозі закономірностей, що втратили чинність.

Відомим виходом з такого стану може служити застосування для прогнозів так званих **адаптивних моделей**. Їхньою особливістю є те, що при

кожному надходженні нової інформації до параметрів моделі вносять відповідні корективи. У результаті модель постійно адаптується до нових умов. При цьому систематично перевіряється близькість її розрахункових даних до фактичних рівнів ряду.

Самою складною проблемою прогнозування, що дотепер не має задовільного розв'язання, є пророкування ламання тенденції, що лежить в основі прогнозних розрахунків. Поки ламання немає, прогнозування дає мінімальні помилки. Коли ж воно відбувається, то самі витончені системи прогнозування дають збій і виникають великі помилки.

Аналітики біржових ситуацій на основі багаторічних спостережень за індексом Доу Джонса стверджують: імовірність ламання тенденцій виникає тоді, коли чергове падіння біржових курсів виявляється більш глибоким, ніж попереднє. Наприклад, якщо динаміка згаданого індексу за чотири періоди має вигляд: 7200, 7160, 7240 і 7180, то падіння, що відбулося наприкінці цього ряду не повинне викликати тривоги. Інша справа, якби на кінці ряду опинилася цифра менше 7160. Це було б сигналом того, що в самому найближчому майбутньому на біржі може відбутися обвал курсів акцій або цінних паперів.

**Стаціонарним динамічним рядом** називається такий динамічний ряд, у якого відсутня тенденція до зростання або падіння, тобто відсутній тренд.

Рівні стаціонарного динамічного ряду коливаються навколо деякого середнього значення. Прогнозування зводиться до пошуку цього середнього значення.

Прогноз для періоду  $t+1$ , тобто на один крок уперед, на основі середнього рівня може мати такий вигляд;

$$P_{t+1} = (\Phi_t + \Phi_{t-1} + \Phi_{t-2}) : 3, \quad (13.24)$$

де  $\Phi_t$  - фактичний рівень поточного періоду;

$\Phi_{t-1}$ ,  $\Phi_{t-2}$  - рівні більш ранніх періодів.

Тут для розрахунку середньої взяті всього три минулих рівні. Їх може бути більше. Погашення випадковості при цьому буде сильнішим. Однак, заглиблюючись в історію, треба завжди пам'ятати про небезпеку перенесення на майбутнє занадто старих закономірностей.

Недоліком вищенаведеної середньої є те, що при її розрахунку надається однакова вага періодам, по-різному віддаленим від горизонту прогнозування. Тим часом періодам, близьким до горизонту прогнозування, треба було б давати більшу вагу. Це якоюсь мірою послабляло б вплив старих закономірностей і посилювало б значення закономірностей останніх років, близьких до горизонту прогнозування. Це особливо актуально для часу великих і різких змін в економіці, що характерно зараз для нашої країни.

Нерівні ваги можна взяти довільно. Наприклад, у такий спосіб:

Період	$t$	$t-1$	$t-3$
Вага	3	2	1

Тоді розрахунок середньої виглядатиме так:

$$\bar{y}_t = \frac{y_t 3 + y_{t-1} 2 + y_{t-2}}{6}$$

Р. Браун запропонував брати ваги, що убують за експонентою:

Період	$t$	$t-1$	$t-2$	$t-3$	...	$t-n$
Вага	$\alpha$	$\alpha(1-\alpha)$	$\alpha(1-\alpha)^2$	$\alpha(1-\alpha)^3$	...	$\alpha(1-\alpha)^n$

Середня з такими вагами називається експонентною.

Величина  $\alpha$  називається параметром згладжування й обчислюється за формулою, запропонованою Р. Брауном:

$$\alpha = \frac{2}{n-1}, \quad (13.25)$$

де  $n$  - число рівнів динамічного ряду, які бажано взяти до уваги при розрахунку середньої.

Якщо прогнозист визнає, що йому треба орієнтуватися тільки на чотири останніх рівні ряду, то:  $\alpha = \frac{2}{4+1} = 0,4$ .

Ваги окремих періодів при цьому матимуть такий вигляд:

Період	$t$	$t-1$	$t-2$	$t-3$	...
Вага	0,4	$0,4(1-0,4)$	$0,4(1-0,4)^2$	$0,4(1-0,4)^3$	...
або	0,4	0,24	0,144	0,0864	...

У зарубіжних роботах із прогнозування пропонується брати  $\alpha$  на рівні 0,05-0,30, тобто орієнтуватися на тривалу історію: 8-40 часових періодів. Це для наших мінливих умов не підходить. Для сучасних умов України параметр  $\alpha$  треба брати на рівні 0,7-0,9. Якщо взяти, наприклад,  $\alpha = 0,7$ , то прогноз для періоду  $t + 1$  буде виглядати так:

$$П_{t+1} = \Phi_t 0,7 + \Phi_{t-1} 0,21 + \Phi_{t-2} 0,063 + \Phi_{t-3} 0,0189 + \dots,$$

«хвіст» цього виразу, тобто

$$\Phi_{t-1} * 0,21 + \Phi_{t-2} 0,063 + \Phi_{t-3} 0,0189 + \dots,$$

можна представити як прогноз для періоду  $t$ . Тоді прогноз для періоду  $t + 1$  має вигляд рекурентної формули:

$$П_{t+1} = \Phi_t \alpha + П_t (1-\alpha). \quad (13.26)$$

Розрахунки кожного чергового прогнозу є при цьому продовженням раніше зроблених розрахунків. Виключається необхідність провадити їх щораз із самого початкового періоду.

При використанні приведеної рекурентної формули для самого раннього періоду вживають «наївний прогноз»  $П_1 = \Phi_1$ .

Можна також замість «наївного прогнозу» використати експертну оцінку. Тоді прогноз для другого періоду на базі даних першого періоду має вигляд:

$$П_2 = \Phi_1 \alpha + ПЕ_1 (1-\alpha), \quad (13.27)$$

де  $ПЕ_1$  - прогнозна величина першого періоду, установлена експертним шляхом.

При експертному оцінюванні величині  $ПЕ_1$  прагнуть надати таке значення, яке б у наступних розрахунках мінімізувало помилку прогнозу. Мінімізація помилки досягається не тільки перебором можливих значень згаданої величини, але також перебором різних значень параметра згладжування  $\alpha$ . Чим менше в кінцевому результаті виходить помилка наступних прогнозів, тим краще буде адаптація моделі до реально сформованих умов.

Коли в змінах рівнів ряду можна доглянути наявність тенденції до зростання або падіння, то говорять, що ряд має тренд, тоді він є **нестационарним**. Наявність тренду повинна мати об'єктивне підтвердження, а не ґрунтуватися на одному тільки візуальному враженні. Для одержання згаданого підтвердження можна скористатися критерієм серій. Відповідно до цього критерію можна говорити про наявність тренду тільки тоді, коли число серій у ряді не перевищує певного критичного значення, узятото з відповідних таблиць. При цьому серією вважається послідовність елементів одного вигляду. Наприклад, послідовність елементів, менших за медіану. Або послідовність елементів, де всі вони більше або дорівнюють медіані.

Розглянемо приклад. Припустимо, що обороти фірми за період із січня по вересень склали, млн.грн.: 10, 12, 11, 13, 13, 14, 13, 15, 13.

Чи можна вважати, що даний ряд дійсно має тренд або у ньому чисто випадкові коливання його рівнів? Для відповіді на це питання підраховуємо число серій, позначаючи рівні, менші за медіану (дорівнює 13) через А. Інші рівні - через В. Дістанемо: АААВВВВВВВ. У наявності 2 серії. Критичне значення, знайдене з таблиць для 3 елементів однієї послідовності, 6 елементів другої і 5%-го рівня значущості перевірки дорівнює теж 2. Перевищення немає. Виходить, гіпотезу про наявність тренду відхилити не можна.

Звісно, може всеж статися, що в даного ряду ніякої тенденції до зростання немає, а вся справа полягає у випадковому коливанні його рівнів. Проте імовірність цього менша за 5%.

За наявності тренду прогноз здійснюють за допомогою регресійних рівнянь. Вони розглядалися вище як моделі динаміки. Найпростішою моделлю є рівняння прямої  $y_t = at + b$ , де  $t$  - фактор часу, тобто порядковий номер рівня ряду.

Підставивши в дане рівняння в якості  $t$  порядкові номери майбутніх періодів часу, одержимо точковий прогноз для цих періодів.

Порівняно із точковим значно більшу практичну цінність має інтервальний прогноз, що має задану імовірність. Імовірність же здійснення точкового прогнозу згідно положень теорії імовірностей дорівнює нулю.

Для одержання інтервального прогнозу попередньо розраховують граничну помилку рівняння. Для її знаходження використовують формулу:

$$\Delta = t_{\alpha}, \quad (13.28)$$

де  $\Delta$  - стандартна помилка (std. error of estimate);

$t$  - квантіль розподілу Стюдента для відповідного числа ступенів свободи, рівного  $n-m$ , і заданої імовірності прогнозу ( $Q$ ), що зазвичай береться на рівні 5%.

Елементарна логіка говорить про те, що помилка прогнозу не може залишатися постійною, незважаючи на збільшення періоду попередження. Вона має зростати. Із збільшенням періоду попередження повинні також розширюватися межі довірчого інтервалу. Інакше кажучи, потрібне виправлення на зміну періоду попередження. Це можна здійснити за допомогою наступної формули:

$$K = \sqrt{\left( \frac{n+1}{n} + \frac{3(n+2L-1)^2}{n^3 - n} \right)},$$

де  $L$  - період попередження;  
 $n$  - довжина ряду.

### 13.6 Зв'язний аналіз часових рядів

Вивчення зв'язку між двома рядами називають **зв'язним аналізом**. До нього прибігають тоді, коли хочуть оцінити ефективність витрат на ті або інші заходи. Наприклад, порівнюють зростання витрат на рекламу з зростанням товарообігу торгового підприємства або динаміку витрат на добрива з динамікою врожайності.

При здійсненні зв'язного аналізу треба завжди пам'ятати про можливість викривлення його результатів за рахунок впливу так званої помилкової кореляції, що може виникнути через просте супуття в часі розвитку двох явищ.

Яскравий приклад помилкової кореляції навів в одній зі своїх робіт англійський статистик Д. Фінні. Він зрівняв динаміку зареєстрованих радіоприймачів і число душевнохворих у післявоєнній Англії і одержав високий коефіцієнт кореляції, хоча прямий зв'язок тут, звичайно, відсутній.

Супуття в часі може виникнути через наявність в часових рядів автокореляції. Під нею розуміють залежність наступних рівнів ряду від попередніх. Для характеристики автокореляції існує тест Дарбіна-Уотсона, що обчислюють за формулою:

$$DW = \frac{\sum_{i=1}^{n-1} (e_{i+1} - e_i)^2}{\sum_{i=1}^n e_i^2}, \quad (13.29)$$

де  $e_i$  - різниця між фактичним і вирівняним рівнем для  $i$ -го періоду.

Автокореляція відсутня і можна без небезпеки викривлень вивчати зв'язок між двома рядами, коли  $DW$  близький до 2. Автокореляція є й можливі великі викривлення рівня зв'язку між рядами за рахунок існування між ними помилкової кореляції, коли  $DW$  близький до нуля або 4.

Існують таблиці значень цього тесту. Входами в них є задана імовірність перевірки і число спостережень  $n$ . Наприклад, для  $n = 15$  і 95%-ї імовірності висновків приводяться такі межі цього тесту:

а) при  $DW < 1,08$  (або  $DW > 2,92$ ) існує позитивна (або негативна) автокореляція;

б) при  $1,08 < DW < 1,36$  і при  $2,64 < DW < 2,92$  існує невизначеність, коли не можна з достатньою впевненістю ні відхилити, ні прийняти гіпотезу про наявність автокореляції;

в) при  $1,36 > DW < 2,64$  автокореляція відсутня.

Наявність автокореляції не тільки утруднює вивчення зв'язку між часовими рядами, але й робить зовсім неможливим здійснення прогнозу значень рівня одного ряду за передбачуваним значенням іншого ряду.

Іноді одразу видно, що ряди ніяк не можуть бути зв'язаними, тому що матеріальна природа відбиваних ними явищ не припускає цього. В інших випадках вирішити питання про наявність або відсутність зв'язку між рядами досить важко. У таких випадках потрібна перевірка рядів на наявність автокореляції і необхідне використання методів, що послабляють або виключають вплив помилкової кореляції.

### **Контрольні запитання**

1. Наведіть визначення часового ряду.
2. Чим відрізняються часові ряди від звичайних просторових вибірок?
3. Які фактори впливають на рівні часового ряду?
4. Які компоненти містить реальний часовий ряд?
5. Поясніть адитивну і мультиплікативну моделі часового ряду.
6. Поясніть, що таке тренд часового ряду і які види тренду зустрічаються?
7. Як визначають значущість тренду?
8. Перелічіть основні етапи аналізу часових рядів.
9. Перелічіть найпоширеніші методи аналізу часових рядів.
10. Поясніть, що таке автокореляція рівнів часового ряду. Як її можна виміряти?
11. Перелічіть властивості коефіцієнта автокореляції.
12. Що таке автокореляційна функція часового ряду і корелограма?
13. З якою метою виконують аналітичне вирівнювання часового ряду?
14. Які функції найчастіше застосовують для побудови трендів?
15. Який метод дозволяє визначити параметри тренду?
16. Які методи використовують для згладжування часового ряду?
17. В чому полягає метод ковзних середніх?
18. Поясніть, як провадиться усунення сезонних коливань за методом ковзної середньої?
19. З якою метою провадиться спектральний аналіз часового ряду?
20. Поясніть, у чому полягає зв'язний аналіз часових рядів?



## СПИСОК ДЖЕРЕЛ

1. Замков О. О. Математические методы в экономике / О. О. Замков. - М.: Финансы и статистика, 2001.
2. Экономико-математические методы и модели : Учебн. пособие / Н. И. Хлод, А. В. Кузнецов, Я. Н. Жихар и др.; Под общ. ред. А. В. Кузнецова. 2-е изд. - Мн.: БГЭУ, 2000. - 412 с.
3. Экономико-математические методы и прикладные модели : Учебн. пособие для вузов / В. В. Федосеев, А. Н. Гармаш, Д. М. Дайитбегов и др.; Под ред. В. В. Федосеева. - М.: ЮНИТИ, 2001. - 391 с.
4. Вітлінський В. В. Математичне програмування: [навч. посіб для студ. вищих навч. закл. економ. спец.] / В. В. Вітлінський, С. І. Наконечний, Т. О. Терещенко. - К.: КНЕУ, 2001. – 380 с.
5. Кузнецов Ю. Н. Математическое программирование: учебник / Ю. Н. Кузнецов, В. А. Кузубов, А. В. Волощенко. - М.: Высш.школа, 1980. – 240 с.
6. Акулич И. Л. Математическое программирование в примерах и задачах : учеб. пособ. / И. Л. Акулич. - М.: Высш. школа, 1986. – 244 с.
7. Таха Х. А. Введение в исследование операций / Х. А. Таха - М.: Изд.дом «Вильямс», 2005. – 912 с.
8. Исследование операций в экономике : Уч. пособие для вузов / Н. Ш. Кремер, Б. А. Путко, И. М. Тришин, М. Н. Фридман./ Под ред. проф. Н. Ш. Кремера. - М.: ЮНИТИ, 2003. - 407 с.
9. Акоф Р. Основы исследования операций / Р. Акоф, М. Сасиени. - М.: Мир, 1971. – 320 с.
10. Наконечный С. И. Эконометрия: учебник / С. И. Наконечный, Т. П. Терещенко. - К.: КНЕУ, 2001.
11. Кремер Н. Ш. Эконометрика : Учебник для вузов / [Н. Ш. Кремер, Б. А. Путко; под ред. проф. Н. Ш. Кремера]. - М.: ЮНИТИ-ДАНА, 2002. - 311 с.
12. Практикум по эконометрике : Учебн. пособие / [И. И. Елисеева, С. В. Курышева, Н. М. Гордиенко и др.; Под ред. И. И. Елисеевой]. - М.: Финансы и статистика, 2002. - 192 с.
13. Магнус Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. - М.: Дело, 2001. - 400 с.
14. Кулинич Е. И. Эконометрия : учеб. пособ. / Е. И. Кулинич - М.: Финансы и статистика, 2000. – 304 с.
15. Афанасьев В. Н. Анализ временных рядов и прогнозирование : Учебник / В. Н. Афанасьев, М. М. Юзбашев. - М.: Финансы и статистика, 2001. - 228 с.
16. Лещинский О. Л. Эконометрия : учеб. пособ. / О. Л. Лещинский, Рязанцева В. В., О. О. Юнькова. - К.: МАУП, 2003.
17. Эконометрика : Учебно-методическое пособие / А. К. Шалабанов, Д. А. Роганов – Казань: ТИСБИ, 2002. – 56 с.
18. Доугерти К. Введение в эконометрику. - М.: Финансы и статистика, 1999. - 402 с.

## ЗМІСТ

ВСТУП.....	3
Змістовий модуль 1. Оптимізаційні економіко-математичні моделі .....	4
Тема 1. Концептуальні аспекти математичного моделювання економіки .....	4
1.1. Економічні моделі. Поняття економічної моделі.....	4
1.2. Принципи моделювання .....	5
1.3. Класифікація моделей .....	7
1.4. Якість моделі.....	9
1.5. Прийняття рішень (вибір).....	9
1.6. Методи прийняття рішень .....	12
Контрольні запитання .....	14
Тема 2. Поняття оптимізаційних задач і оптимізаційних моделей.	
Класифікація методів .....	15
2.1 Основні поняття оптимізаційних задач і моделей .....	15
2.2 Класифікація методів оптимізації.....	18
Контрольні запитання .....	19
Тема 3. Лінійне програмування .....	20
3.1. Загальна форма задачі лінійного програмування (ЗЛП) .....	20
3.2. Основні властивості ЗЛП і її перша геометрична інтерпретація .....	21
3.3. Канонічна форма задачі лінійного програмування (КЗЛП).....	25
3.4. Симплекс-метод.....	28
Контрольні запитання .....	35
Тема 4. Теорія двоїстості і двоїсті оцінки в аналізі розв'язків ліній- них оптимізаційних моделей.....	36
4.1. Пряма і двоїста задачі як пара сполучених задач ЛП.....	36
4.2. Основні теореми двоїстості, їхній економічний зміст.....	38
4.3. Двоїсті оцінки і дефіцитність ресурсів.....	41
Контрольні запитання .....	41
Тема 5. Аналіз лінійних моделей економічних задач.....	42
5.1. Аналіз розв'язків лінійних економіко-математичних моделей .....	42
5.2. Аналіз параметричної стійкості розв'язків ЗЛП .....	46
5.3. Оцінка рентабельності виробленої продукції.....	48
5.4. Аналіз обмежень дефіцитних і недефіцитних ресурсів.....	49
Контрольні запитання .....	49
Тема 6. Транспортна задача.....	50
6.1. Транспортна задача в матричній постановці і її властивості.....	50
6.2. Методи побудови опорного плану.....	52
6.3. Метод потенціалів .....	55
6.4. Випадок виродження.....	58
6.5. Транспортна задача за критерієм часу .....	60
Контрольні запитання .....	61
Тема 7. Цілочислові задачі лінійного програмування. Основні методи їх розв'язання і аналізу .....	62
7.1. Типи задач дискретного програмування .....	62
7.2. Метод Гоморі .....	65

7.3. Метод віток і границь.....	67
Контрольні запитання .....	69
Тема 8. Задачі нелінійного програмування. Основні методи їх розв'язання і аналізу .....	70
8.1. Постановка задачі нелінійного програмування (ЗНП) .....	70
8.2. Класичний метод оптимізації з використанням множників Лагранжа. 72	
8.3. Опукле програмування.....	73
8.4. Необхідні й достатні умови існування сідлової точки. Теорема Куна-Таккера.....	76
8.5. Деякі методи розв'язання задач НЛП.....	79
Контрольні запитання .....	84
Змістовий модуль 2. Економетричні моделі.....	85
Тема 9. Принципи побудови економетричних моделей.....	85
9.1. Роль економетричних досліджень в економіці .....	85
9.2. Етапи економетричного моделювання.....	86
9.3. Класифікація економетричних моделей.....	87
Контрольні запитання .....	92
Тема 10. Методи побудови загальної лінійної моделі.....	92
10.1. Побудова загальної лінійної моделі .....	92
10.2. Лінійна модель парної регресії. Суть методу найменших квадратів.. 96	
10.3. Оцінка значущості рівняння лінійної регресії та перевірка моделі на адекватність за критеріями Стюдента і Фішера.....	100
10.4. Лінійна модель множинної регресії.....	103
10.5. Оцінка значущості множинної регресії і показники якості моделі... 107	
Контрольні запитання .....	113
Тема 11. Мультиколінеарність і її вплив на оцінки параметрів моделі .....	114
11.1. Поняття мультиколінеарності. Вплив мультиколінеарності на оцінки параметрів .....	114
11.2. Методи виключення мультиколінеарності .....	116
Контрольні запитання .....	120
Тема 12. Узагальнений метод найменших квадратів .....	120
12.1. Поняття гомоскедастичності і гетероскедастичності .....	120
12.2. Узагальнений метод найменших квадратів (УМНК).....	125
Контрольні запитання .....	129
Тема 13. Економетричні моделі динаміки .....	129
13.1. Загальні відомості про часові ряди і завдання їхнього аналізу .....	129
13.2. Моделювання тенденції часового ряду .....	133
13.3. Аналіз аддитивної і мультиплікативної моделей часового ряду.....	135
13.4. Спектральний аналіз часового ряду .....	138
13.5. Прогнозування часового ряду .....	139
13.6. Зв'язний аналіз часових рядів.....	142
Контрольні запитання .....	143
Список джерел .....	145

*Навчальне видання*

**ВОРОНКОВ** Олексій Олександрович

КОНСПЕКТ ЛЕКЦІЙ  
з курсу

# **ЕКОНОМІКО-МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ**

*(для студентів заочної форми навчання освітньо-кваліфікаційного рівня  
«бакалавр» напрямку підготовки 6.030601 - Менеджмент)*

Відповідальний за випуск *Т. А. Пушкар*

За авторською редакцією

Комп'ютерне верстання *І. В. Волосожарова*

План 2015, поз. 183 Л

---

Підп. до друку 02.07.2015

Формат 60x84/16

Друк на ризографі

Ум. друк. арк. 8,7

Тираж 50 пр.

Зам. №

Видавець і виготовлювач:

Харківський національний університет  
міського господарства імені О. М. Бекетова,  
вул. Революції, 12, Харків, 61002

Електронна адреса: [rectorat@kname.edu.ua](mailto:rectorat@kname.edu.ua)

Свідоцтво суб'єкта видавничої справи:

ДК №4705 від 28.03.2014 р.